NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

How to effectively use interactivity to improve visual analysis and communication in groups of novices or experts

Maria do Mar Guiod de Castro Couto Viana

Project Work presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

How to effectively use interactivity to improve visual analysis and communication in groups of novices or experts

by

Maria do Mar Guiod de Castro Couto Viana

Project Work presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Knowledge Management and Business Intelligence

Advisor: Pedro da Costa Brito Cabral

May 2020

DEDICATION

Esta tese não teria sido possível sem a ajuda fundamental de várias pessoas na minha vida às quais estou extremamente grata.

Ao meu orientador, Pedro Cabral, por todo o apoio e ajuda que meu deu ao longo deste percurso.

A todas as pessoas que participaram neste estudo e me ofereceram um bocadinho do seu tempo para me ajudar, em especial ao Vítor Silveira, que sempre me ajudou desde o início e foi incansável para sempre conseguir apoiar-me de todas as maneiras possíveis.

Aos meus pais, por todo o apoio que sempre me deram, por me poderem ter dado esta educação e ajudarem-me sempre a concretizar todos os meus objetivos.

Ao meu namorado, por estar sempre lá nos bons e nos maus momentos, por ter tido a paciência de me aturar mesmo nas alturas mais difíceis e por mostrar sempre o quanto me apoiava.

A toda a família e amigos que me acompanharam nesta etapa e que estiveram sempre ao meu lado e se mostraram interessados neste meu projeto.

Muito obrigado a todos, sem a vossa ajuda crucial nada disto seria possível.

RESUMO

Conseguir passar um ponto de vista claro através da visualização de dados é um dos principais objetivos das organizações dos dias de hoje. O principal objetivo deste projeto foi perceber qual a melhor maneira de utilizar interatividade em diferentes tipos de indivíduos, especialistas e novatos, e descobrir as principais diferenças entre os dois grupos. Isto foi feito através da criação de um protótipo com diversas visualizações interativas, onde em cada uma delas foram utilizadas diferentes técnicas de visualização e interatividade. Após a sua criação, seguiu-se a validação de cada uma delas, de modo a chegar a conclusões sobre os melhores métodos a utilizar para melhorar a análise e comunicação da informação, para os diferentes grupos de indivíduos. O desenvolvimento do protótipo foi realizado com o software R, mais especificamente o pacote Shiny. O estudo contribuiu com uma metodologia para avaliar as diferenças entre grupos de especialistas e de novatos, relativamente ao protótipo de visualização que foi validado recorrendo a 6 medidas quantitativas e qualitativas. Utilizando um teste ANOVA de fator único foi possível concluir que em relação às medidas quantitativas não foram encontradas diferenças com significância estatística e em relação às medidas qualitativas a única medida que mostrou diferenças com significância estatística entre ambos os grupos foi o nível de interação (*engagement*). Isto significa que esta é a única métrica possível de melhorar para diminuir as diferenças entre ambos os grupos. Em relação às visualizações ambos os grupos, concordaram que as melhores foram o mapa de calor (*heatmap)* e o gráfico de barras e as piores visualizações foram o mapa coropleto e o gráfico de barras empilhadas. Houve, no entanto, diferenças entre a forma como os diferentes grupos interagiram com os componentes. Por exemplo, a *select box* foi uma melhor opção para o grupo de novatos, enquanto que a *radio box* foi a melhor para o grupo de especialistas. Os *tooltips* e o *slider* foram adequados para os dois tipos de indivíduos. Também foi comprovado que o pacote Shiny é uma ferramenta capaz de criar visualizações interativas eficazes para diferentes tipos de indivíduos uma vez que, em média, os participantes obtiveram ótimos resultados utilizando medidas qualitativas ou quantitativas. Os resultados deste estudo, permitirão às organizações a adaptação eficiente das suas visualizações a diferentes tipos de audiência.

PALAVRAS CHAVE

Visualização de dados, Interatividade, Especialistas, Novatos, R, Shiny

ABSTRACT

Getting a clear point of view through data visualization is one of the main goals of todays' organizations. The main objective of this project was to understand the most efficient way to use interactivity in different groups of individuals, experts and novices, and to discover the main differences between these two groups. This was achieved through the creation of a prototype with several interactive visualizations, where in each of them different visualizations and interaction techniques were used. After the creation of the prototype, the next step was the validation of each one of them to reach conclusions on what are the most effective means to improve visual analysis and communication, in different groups of individuals. The development of the prototype was done using the R software, and most specifically the Shiny package. This study contributed with a methodology to evaluate the differences between experts and novices, using the visualization prototype that was validated with 6 quantitative and qualitative metrics. Using an ANOVA single factor test it was possible to conclude that regarding the quantitative measures no statistically significant differences were found. However, regarding the qualitative measures the only measure that had statistically significant differences between both groups was the engagement measure. This means that this is the only metric where results can be improved in order to close the gap between the group of experts and novices. Regarding the visualizations, both groups agreed that the best visualizations were the heatmap and the bar chart and the worst visualizations were the choropleth map and the stacked bar chart. Nevertheless, there were differences between how the different groups interacted with the components. For example, the select box was a better option for the novice's group, while the radio box was the best for the expert's group. The tooltips and the slider are adequate for both types of individuals. It was also proved that the Shiny package is a tool that is capable of creating effective interactive visualizations for different types of individuals, since that on average the participants obtained great results using qualitative or quantitative measures. The results of this study will allow organizations to efficiently adapt their visualizations to different types of audiences.

KEYWORDS

Data visualization, Interactivity, Experts, Novices, R, Shiny

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

GDP Gross domestic product

SI Suggested interactivity

1. INTRODUCTION

1.1. BACKGROUND AND PROBLEM IDENTIFICATION

To start this study, it is first necessary to identify the business problem at hand, which is: How can organizations effectively use interactivity and which visualizations techniques can they use, to build interactive data visualizations, that convey the information in the most effective manner to different groups of individuals?

This problem can have relevance in the business world considering that data visualization is an increasingly used tool to help users in the decision-making process (Diamond & Mattia, 2017). This could possibly mean that if an organization uses the best visualization and interaction techniques, combined with different tools and algorithms, they may be able to create interactive visualizations that improve visual analysis and communication, in distinct types of individuals.

First, it is critical to realize the importance that interactivity has in visual analysis. It has been discovered that interactivity is a very effective instrument to use when creating a visualization since it creates a better comprehension and helps with communication (Alexandre, 2016). Moreover, using novel algorithms and encodings with tools such as the R package Shiny, it is possible to create interactive web base applications, making it possible to easily integrate interactivity in the visualizations created (Ellis & Merdian, 2015).

All the arguments presented above have important relevance in the reason of why this study is important to conduct since it is proof of how effective the use of the tool in mention and interactivity can be in visual analysis and communication. Therefore, using them together and combining the best visualization techniques could present the organizations with very effective interactive visualizations that could be extremely useful in the business world, especially when targeting a general audience.

This study provides a broader view on how organizations should use interactivity to better pass the message they want to share, to different groups of individuals and which visualizations techniques could be more effective with the use of interactivity.

1.2. STUDY RELEVANCE AND IMPORTANCE

Although there are some studies that tried to find differences between how different types of individuals discover interactivity (Blascheck, et al, 2019; Oghbaie, Pennock, & Rouse, 2016; Rouse, Pennock, Oghbaie, & Liu, 2017), none of these studies used an interactive visualization related with the field of economy. Furthermore, none of these studies, had the specific goal of testing which were the best visualization idioms for each type of group.

Additionally, although some publications prove the benefit of interactivity in visual analysis and communication (Alexandre, 2016; Blascheck, et al., 2019; Mottus, Kinshuk, Graf, & Chen, 2013; Dove & Jones, 2012), not so many can be found using R statistical environment and more specifically, Shiny apps. Usually, these publications are very specific and related to an actual application built using Shiny, to use in a very specific field. Moreover, none of the studies that attempted to find disparities in how the various sorts of individuals find interactivity, have used this type of tool.

One type of organization that highly benefits with the correct use of data visualization is one that works in journalism (Alexandre, 2016) or any organization that wants to tell a story with data. Data visualization is crucial in journalism since it is a highly effective method to tell complex stories but also is a way to secure the liability of journalism nowadays (Alexandre, 2016). Furthermore, interactivity can also help users gain insights and improve decision making (Alexandre, 2016). Therefore, any company or organization that works in the journalism area would benefit from this study, since it would find other useful methods and tools to create visualizations that are able to effectively tell a story to a general audience, while still using interactivity.

Additionally, any company that has a platform accessible to a general audience, and that uses data visualization to improve that platform, could benefit from this work. More specific examples could be, for instance, organizations that work in education (Mottus, Kinshuk, Graf, & Chen, 2013) or in finance (Zhang, 2015) . This shows that data visualization is important in a broad set of fields. Therefore, it could be interesting to analyse how different types of individuals discover interactivity and which are the best visualization techniques for each group, since companies and organizations can create visualizations that will be exposed to different types of audience.

In this study, we built a prototype that uses data from the National Bank of Portugal, more specifically data regarding the balance of the payments, to create an interactive visualization that could be used in a website to better explain specific information to a general audience.

The fact that the Portuguese economy requested in 2011 a Financial and Assistance Programme from the International Monetary Fund, European Commission and European Central Bank, motivated by external imbalances, encouraged using this topic in the building of this prototype. The current and capital account statistics measure was used, reflecting, monthly, net lending/net borrowing of the Portuguese economy vis-à-vis the rest of the world.

1.3. OBJECTIVES

The research questions of this study are the following:

- Are there differences in how the different types of individuals (experts or novices) discover interactivity?

- Which are the best visualization and interaction techniques an organization should use in their visualizations, to disseminate information to different types of individuals?

- Using the R package Shiny open tool, is it possible to create visualizations that conjugate interactivity to improve visual analysis and communication in different groups of individuals, experts or novices?

Data is beginning to be more and more accessible to the public and is, therefore, important to realize what exactly is the best way for not only the experts but also the general audience to understand a visualization (Blascheck, et al., 2019).

Currently, there is a gap in how general individuals understand interactive functionality (Blascheck, et al., 2019), so the main study objective of this project is to discover the most efficient way to use interactivity and which are the best visualization techniques to apply, while using open software tools such as the R package Shiny, in order to find the most efficient way to disseminate information either to groups of experts or to groups of novices and simultaneously discovering the differences between these two groups. This can be related with the business problem at hand, considering that a main goal is to help

organizations in different fields, create interactive visualizations with the most effective visualization techniques for different types of individuals, using the tool mentioned above.

Consequently, a specific goal of this project is to create a prototype, using the tool mentioned above, with several different interactive visualizations, using distinctive visualization and interaction techniques in each of them, to understand which are the most effective in improving visual analysis and communication for the different types of individuals.

After building the initial prototype, another specific objective of this study is also to find the best methods to validate it. The feedback collected from this validation will then be used to continue to improve the prototype and to try to increasingly reach the main goal of this study.

1.4. THESIS STRUCTURE

This project will be structured in 5 chapters.

In the first chapter, an introduction for the project will be done, explaining the background and the scope of this study, to better understand the project goals.

In the second chapter, theoretical background research will be done, as well as the continuation of the literature review, to fully justify the project being done and how is it going to possibly contribute to the state of the art.

In the third chapter, all the methodologies and datasets used to develop the prototype will be explained as well as the justification of the choices used. In this chapter, it will also be explained the process of how each one of the interactive views of the prototype were evaluated, stating all the validation types used and how the process was conducted.

In the fourth chapter, the results collected from this study will be reported as well as all the improvements that were done to the prototype.

In the fifth chapter, the discussion of the results will be made. The goal of this chapter is to criticize the data collected from the evaluation phase and compare it to the state of the art.

Finally, in the last chapter, the conclusions reached from this study will be explained, as well as the limitations of the project and the future work other researchers may do, applying the knowledge of this study.

2. LITERATURE REVIEW

This chapter provides with an overview of what has already been done regarding topics related to this project and is divided into four sections: discover of interactivity in different groups of subjects, suggested interactivity, tools to develop interactive visualizations and validation of interactive visualizations. In each section, it will be discussed the different approaches that have been used in past studies and the corresponding strengths and weaknesses of each method.

2.1. DISCOVERY OF INTERACTIVITY IN DIFFERENT GROUPS OF SUBJECTS

Previous studies have tried to discover how different groups of subjects discover interactivity, but in a specific study, no differences between these groups have been found (Blascheck, et al., 2019). These authors used eye-tracking technology and three types of individuals (novices, savvy, and experts), to understand how they discovered interactivity. In this work, the participants were exposed to an interactive visualization and were asked to explore it, while using eye-tracking technology to document the process. Although it was not possible to find differences between each group of individuals, it was possible to identify different exploration strategies that users, in general, may apply when interacting with interactive visualizations.

Using the different exploration strategies it was possible to create six suggestions. These suggestions are recommended to use when creating an interactive visualization (Blascheck, et al., 2019):

- **Inviting interaction**: which means using animations or emphasis to invite the user to interact, since it is known that users may take a significant amount of time to start interacting;

- **Combating oscillation**, thus users don't have to go back and forth between the views to understand an information since this could create annoyance in the viewers;

- **Leveraging spatial organization**, subsequently simplifying the organization of the controls;

- **Providing entry points**, such as using information that would be familiar to the users when using the interactive visualizations and that would, therefore, increase the engagement with it;

- **Scaffolding complex interactions**, such as giving hints or tutorials on how to use a more complex interaction to help users use the visualizations to the maximum of their potential;

- **Supporting transitions**, which means implementing smoother transitions between views and between datasets in the interactive visualizations.

Using these suggestions it should be possible to improve the functionality of interactive visualizations in different groups of individuals (Blascheck, et al., 2019).

In the experiment of the work conducted by Blascheck, *et al.* (2019), when interacting with the visualization, participants did not have any particular task to solve and they were only told to freely explore the visualizations. This also proved to be a limitation since it was not possible to infer if the user's actions were intentional or not.

Other works have obtained different conclusions regarding the differences in how distinct groups of individuals interact with visualizations, such as the work conducted by Oghbaie, Pennock, & Rouse (2016), that used two different types of individuals, experts, and non-experts.

Although interactive visualizations have become an increasingly prominent method to help users in decision making, it is not sure whether or not they actually improve the decisions (Oghbaie, Pennock, & Rouse, 2016). Therefore, it was necessary to test if by increasing the intricacy of a problem and by using visualization methods that are not appropriate, this will lead to decision-makers identifying misleading correlations as casual relationships. In the work conducted by Oghbaie, Pennock, & Rouse (2016), it was suggested that if the proper visualization methods are applied these effects can be mitigated and the gap between experts and non-experts could be closed. To test this hypothesis an experiment was conducted. The conclusions of this work can be summarized in **Table 1**.

In a similar work conducted by Rouse, Pennock, Oghbaie, & Liu (2017), using the Rasmussen's abstraction-aggregation hierarchy methodology and the same enterprise problem

of the work conducted by Oghbaie, Pennock, & Rouse (2016), it was possible to reach similar conclusions.

The Rasmussen's abstraction-aggregation hierarchy (Rouse, Pennock, Oghbaie, & Liu, 2017) is a methodology that starts by identifying information use cases and then finds which levels of abstraction and aggregation are necessary for the information present on the use cases. In the following step, it takes this information and designs the visualizations and controls. Afterward it integrates the visualization and the controls and lastly integrates across the use cases, to decrease the total number of visualizations and controls.

Using this methodology and the two groups of individuals, experts and non-experts, same as the work conducted by Oghbaie, Pennock, & Rouse (2016), another experiment was conducted using two quantitative measures, speed and accuracy (Rouse, Pennock, Oghbaie, & Liu, 2017). The results of this experiment can also be summarized in **Table 1**.

Table 1 - Differences and similarities between the different groups of individuals

Metrics / Author	**Oghbaie, Pennock, & Rouse (2016)**	**Rouse, Pennock, Oghbaie, & Liu (2017)**
Accuracy	Experts only exceed non-experts in the data with the most complex casual relationships	Experts were more accurate than non-experts
Speed	Same speed for both groups of experts and non-experts	
Data	Experts made full use of all the information displayed whereas non-experts demonstrated a lack of information seeking behaviour	

As for future work, it was suggested that providing training and aiding may affect the results and enable the non-experts group to reach higher results than the experts group. Additionally, it is also suggested to pick a different group of individuals for the group of the non-experts, that still had no major knowledge on the subject but that showed more interest in it, in order to increase the engagement with the visualization (Rouse, Pennock, Oghbaie, & Liu, 2017).

2.2. SUGGESTED INTERACTIVITY

Understanding how to use interactivity in visualizations is crucial to develop techniques that increase the intuitiveness of online visualizations. Suggested Interactivity (SI) can be defined as: “a set of methods for indicating that a graphical area can be interacted with by subtly directing a user’s attention so as not to impede too heavily on this person’s focus or on the rest of the interface design” (Boy, Eveillard, Detienne, & Fekete, 2016).

The work conducted by Boy, Eveillard, Detienne, & Fekete (2016) relies on the use of SI cues and takes important conclusions regarding this topic. This work had three stages. The first one was the execution of three experiments to survey the necessities of these methods. The second one was the creation of a design space with SI cues embedded in the text and the final one the validation of three SI cues used, applied to bar charts.

According to these authors there are three types of SI cues that can be applied in visualizations:

- SI cues that are present in the object of interest, for example in the visualization itself;
- SI cues that are present in external objects, for example in widgets;
- SI cues that use a mix of the first two SI cues and are therefore present in the object of interest and in external objects. These SI cues can also provide feedforward.

The main conclusions of this work were that the vast majority of individuals don’t have the initial motivation to start interacting with visualizations when these are embedded with text and that of the three SI cues tested, the third one, that uses a mix of both the object of interest and an external object, while providing feedforward, is the cue that tempts more users to interact (Boy, Eveillard, Detienne, & Fekete, 2016).

Lastly, for future work, this article recommended that these SI cues should also be tested on different tasks as well as using more complex visualizations that are not just embedded with text.

Additionally, the work conducted by Sun & Hsu (2012) also took important conclusions on how interactivity should be used. In this case, interactivity was not tested in a specific visualization but on websites. Three websites with three different levels of

interactivity (low, medium and high) were tested. In the experiment of this work, the participants were also told to answer a questionnaire related to each one of websites tested.

It has been discovered that websites with a higher interactivity level are the most useful ones to learn. However, it was also found that websites with fewer interactivity functions are usually preferred even though the number of interactivity functions does not directly correlate with the level of interactivity. Another interesting conclusion was the fact that the interactivity level that produced the highest scores in the questionnaire performed was the medium level. In addition, when creating a website, the goal should be to align enough interactive functions with a high level of interactivity. Nevertheless, it was also discovered that a higher level of interactivity may not assure a higher perception of interactivity (Sun & Hsu, 2012).

Sun & Hsu (2012) also developed a few suggestions to take in mind when creating a website. The website should have:

- Enough interactivity functions, that the users will need to use to achieve their tasks;
- Organized content with complete diagrams, tables, and multimedia;
- The possibility to have a real-time discussion;
- An interface that is perceived as easy to use and facilitates navigation.

Using these suggestions, it should be possible to achieve a suitable interactivity level in the visualizations or websites created.

2.3. TOOLS TO CREATE INTERACTIVE VISUALIZATIONS

Nowadays, there are several different tools that can be used to create interactive visualizations. We can distinguish them between ready-to-use tools and tools that require programming skills.

One example of these ready-to-use tools is Power BI, which is a business analytics service provided by Microsoft (Microsoft Power BI, 2020). Power BI offers a set of tools to help the users with the manipulation, analysis and visualization of data, which in turn, makes it possible to create visualizations.

Power BI also makes it possible to perform data cleaning, statistics, and interactive graphics. Although there are other tools that surpass Power BI in these specific tasks, with Power BI it is feasible to perform all these different tasks in only one tool while providing a user-friendly interface at the same time (Becker & Gould, 2019).

There are some studies that used Power BI to develop an interactive visualization for research, for example, the work conducted by Viorel & Lucia (2019), that used Power BI to analyse tourism in the European Union. This study concluded that Power BI is an efficient tool to create interactive data visualizations and reports, and that makes it possible to easily add, modify and visualize the data available.

Another ready-to-use tool is Tableau Desktop, which can also be utilized to create interactive visualizations since it is a data visualization tool that helps the discovery of valuable insights from the data available at a very high speed, and potentials the creation of interactive dashboards or reports. Bhardwaj & Baliyan (2019) mentioned that a few good reasons to use Tableau to create a visualization tool are:

- It has a smaller learning curve than some programming languages like R or Python, and requires less technical expertise;

- It is not as expensive as other tools such as OBE by Oracle or Business objects by SAP.

One example of a tool that requires programming skills to create visualizations is Python, which is a general-purpose programming language that can also be applied in data visualization. In the work conducted by Fahad & Yahya (2018) it was mentioned that Python provides different libraries for data visualization, for example, Bokeh, Seaborn, Altair, ggplot, plotly, and Pygal. Python also provides different tools to help in Big Data Visualization as well (Fahad & Yahya, 2018). Lastly, Python also provides with a framework designed to help create online applications, named Dash, which is also open source.

An alternative example can be the D3.js JavaScript library, which makes it possible to create interactive data visualizations in web browsers. Although it is a highly quality tool to develop customizable data visualizations, for data analysis, other tools such as Python or R can be a better choice due to their extensive libraries options, when compared to JavaScript.

Finally, another programming tool that could be used to create interactive visualizations is the R programming language and software, which also has some of the same libraries as Python, such as ggplot and plotly and also has the possibility to use the Shiny package. Shiny is an R package that enables the creation of interactive visualizations that could be displayed online (Ellis & Merdian, 2015). Shiny uses a user interface-server architecture and has a wide range of widgets that could be used to perform interactions on the visualizations, for example, radio boxes, sliders or select boxes. Besides that, it is also capable of developing interactive plots and different layouts to increase engagement with its visualizations.

As mentioned previously, the research conducted in the introduction of this study, led to believe that not many studies mention the use of the R software, and more specifically the Shiny package. The bibliography found was usually related with applications about highly specific subjects such as analysis of RNA-seq count data (Su, Sun, Shimizu, & Kadota, 2019) or for instance drug target network analysis (Seal & Wild, 2018). One exception was the work conducted by López *et al.* (2018) that through the help of focus groups and questionnaires, tested different Shiny applications with a group of students, to try to reach conclusions about how the students regarded this tool.

Through this work, it was possible to discover several conclusions regarding the opinion of the students. One of these conclusions is the fact that the interfaces of the Shiny applications are intuitive and easy to use. Additionally, it was also discovered that simpler applications are preferred, instead of applications with too much concentrated information in one place (López et al., 2018). Moreover, it was also found that applications created using the Shiny package are useful and help complement the studies of the topics found in them. It is also more appreciated when data is shown in more than one way, although it was discovered that if there are more than two graphics presented in a visualization, usually not all of them are used. Additionally, it is also suggested to create guidance for the applications, such as tutorials or examples to help with their initial analysis (López et al., 2018).

López *et al.* (2018) also developed the following suggestions to be applied when creating a Shiny application:

- The application should be developed taking in mind its main goals in order to prioritize the most important information;

- The application should be clear and intuitive and use different visualization and interaction methods;

- The participants that use the application should have an active attitude when interacting with the visualization.

When using these suggestions, it should be possible to create an efficient Shiny application that increases the engagement with the user and improves the visual analysis and communication with the individuals that interact with it. Therefore, these suggestions will be applied in the prototype of this study.

2.4. VALIDATION OF INTERACTIVE VISUALIZATIONS

Regarding the topic of validation of interactive visualizations, it can be interesting to analyse the framework created by Munzner (2015) to design visualizations. That framework consists of four nested levels:

- **Domain situation**: in this top-level it is necessary to understand which target users and which requirements are necessary for the visualization;

- **Data/Task abstraction**: in this level, the crucial goal is to discover which data will be used in the visualization;

- **Visual encoding/Interaction idiom:** in this level, the objective is to realize what types of visual encodings or interactions will be used to display the data in the visualization;

- **Algorithm:** finally, at the bottom level is necessary to realize if the computer and the code created are effective at displaying the visualization.

Applying this framework can help in the validation process considering that when validating a visualization all these four levels should be analysed. Regarding the top level, the main concern should be if the requirements of the project are being fulfilled. Continuing with the next level, the goal when validating a visualization should be to ensure that the data that is being visualized is the correct data considering the requirements of the visualization.

Concerning the third level, another way to validate the visualization is to make sure that all visualization and interaction idioms applied are the correct ones to display the type of data being used. Finally, for the bottom level, an easy way to validate the visualization is to discover if the code and tool being used to show the visualization are working correctly.

Munzner (2015) provides several validation examples used in different types of works (Heer, Kong, & Agrawala, 2009; Henry & Fekete, 2006; Mcguffin & Balakrishnan, 2005; Mclachlan, Munzner, Koutsofios, & North, 2008; Noack, 2003; Phan, Xiao, Yeh, Hanrahan, & Winograd, 2005), applying the same four-level framework, although not all different levels were applied in the examples presented. The summary of the different types of validation examples and the types of visualizations applied can be seen in **Table 2**.

Table 2 - Different types of validation/visualization applied (adapted from Munzner, 2015)

Author	Type of visualization	Type of validation
(Heer, Kong, & Agrawala, 2009)	Genealogical graphs	Justification of the design applied Qualitative analysis Test of the prototype on users to collect data
(Henry & Fekete, 2006)	Node–link and matrix representations	Observation of the selected users and interviews Justification of the design applied Measurement of the system time/memory Qualitative analysis
(Mcguffin & Balakrishnan, 2005)	Flow map	Justification of the design applied Computation complexity analysis Measure system time/memory Qualitative analysis
(Mclachlan, Munzner, Koutsofios, & North, 2008)	Time-series data with a reorderable matrix	Observation of the selected users and interviews Justification of the design applied Qualitative analysis Field study, documentation of the use of the visualization
(Noack, 2003)	Node–link graphs	Qualitative and quantitative analysis
(Phan, Xiao, Yeh, Hanrahan, & Winograd, 2005)	Line charts, one-band horizon graphs, and two band horizon graphs	Measurement of human time/errors with a lab study

As seen in the examples previously mentioned, an efficient way to validate an interactive visualization can be through the use of quantitative measures such as questionnaires or surveys and/or through the use of qualitative measures such as interviews (Lu, et al., 2018).

This strategy was applied in the work conducted by Lu, *et al.* (2018), where an experiment was performed to test an interactive visualization. In the experiment, the participants had time to first get acquainted with the interactive visualization and afterward they were asked to finish a set of tasks, to determine if the visualization was effective to gather the information needed, in a short period of time. This particular part of the experiment could solve a limitation in this field since instead of just freely exploring the interactive visualization (Blascheck, et al., 2019), the participants had to follow instructions and answer questions, making their actions intentional and not just random. Furthermore, the participants were also interviewed and had to answer a 5 point Likert scale questionnaire about their feelings and judgments on the visualization they tested, documenting also their engagement level with the visualization.

Additionally, there are other works that also discuss the use of quantitative and qualitative measures to measure the efficacy of a visual analytic tool in order to improve data comprehension. An example of this is the work conducted by Géryk (2015), that carried out an experiment, where the participants had to complete a set of tasks, with two datasets of different sizes and an interactive visualization.

Through this work, it was possible to discover that when using these quantitative and qualitative measures the results of an experiment can be divided into three sections: accuracy, completion time and the subject's preferences. The accuracy can be measured by using the percentage of the correct answers of the tasks performed, that the subjects of the experiment may have to complete. The completion time can be measured by measuring the time it takes the subjects of the experiment, to complete the tasks asked. Finally, the subject preferences can be measured by asking the subjects to answer a questionnaire using a Likehert scale, that measures different qualitative measures such as the helpfulness and the interest of the visualization (Géryk, 2015).

When using these quantitative and qualitative measures, Géryk (2015) concluded that, when it comes to large datasets, animated methods lead to fewer errors when compared to

static methods. Géryk (2015) also discovered that animated methods are more efficient in holding the user's attention, possibly due to the efficacy of the animation. This also means that users can feel more entertained when testing with interactive methods and it may increase their interest.

Finally, another two examples that have also used tasks to validate an interactive visualization have been the work conducted by Paul *et al.* (2019), where the participants, that in this case had expertise about the subject of analysis, were asked to complete three tasks using the interactive visualization available, in order to understand if they could use the visualization effectively and the work conducted by Pena-Araya, Pietriga, & Bezerianos (2019), where the participants of the experiment had to complete a sequence of tasks where the goal was to recognize if two variables had a high correlation.

2.5. CONCLUSION

This chapter provided a starting point for the development of the experiment of this study.

For instance, considering this study tried to find the most effective ways to use interactivity in two different groups of individuals, one possible contribution to the state of art might be if it discovers differences between these two groups, since in the work conducted by Blascheck, *et al.* (2019) it was not possible to find any differences between them.

On the other hand, it could also add value if it found different conclusions than the work conducted by Oghbaie, Pennock, & Rouse (2016), and Rouse, Pennock, Oghbaie, & Liu (2017) or even if it supported their work or found further conclusions. The suggested interactivity techniques were also used in the prototype and that could also add value to this study considering it could be interesting to test these SI cues with different and more complex visualizations than just visualizations embedded with text (Boy, Eveillard, Detienne, & Fekete, 2016).

Through this research, it was possible to discover several types of tools that can create interactive visualizations and choose the one that created the prototype of this study, which was the R software and most specifically the Shiny package. Furthermore, considering this fact, this study can possibly also add value to the research being conducted, since none of the previous works stated in this research, used this type of software to develop their visualizations, or they used it to simply developed a very specific application (Seal & Wild,

2018; Su, Sun, Shimizu, & Kadota, 2019), without the goal of testing the interactivity or the visualization methods of the application. The only exception was the work conducted by López *et al.* (2018) that assessed the efficacy of the Shiny applications with students and using the suggestions presented in this work it will likely be possible to create an application that follows all the requirements and increases the engagement with the user, therefore using interactivity and other visualization techniques to improve the visual analysis and communication. Besides this, another way where this study can possibly add value to the research already developed is the fact that this application will be tested on different types of individuals on the contrary of the applications used in the work done by López *et al.* (2018).

Regarding the validation of the interactive visualizations, this study may also possibly complement the previous research by using quantitative and qualitative measures (Géryk, 2015; Lu, et al., 2018) to validate the distinct visualizations that compose the prototype and to test it on different types of individuals, as well as using the validation approach designed by Munzner (2015) to validate the prototype of this study.

Through this research, it was possible to analyse all the different methods, suggestions and tools that have been used in past researches and analyse the advantages and disadvantages of each of them to discover which ones were the most useful to use in the experiment of this project.

Finally, it was also possible to discover previous works that have obtained conclusions regarding topics that are related to this project, and using those conclusions, in the next chapters it was possible to compare the results of the experiment of this study with their results.

In the next sections, using the arguments, the different points of view and strategies presented in the above literature review it was possible to try to find a strategy that tried to solve the limitations that other researchers had in their studies and possibly complement their studies.

3. DATA AND METHODS

In this section, it will be discussed all the methodologies and frameworks used to build the prototype, the justification of the choices applied in each one of the components of the prototype, the data sources utilized, and the methods used for collecting the validation data.

The methodology flowchart used for this experiment took inspiration in the design and creation process suggested by Oates (2005) and can be seen in **Figure 1**.

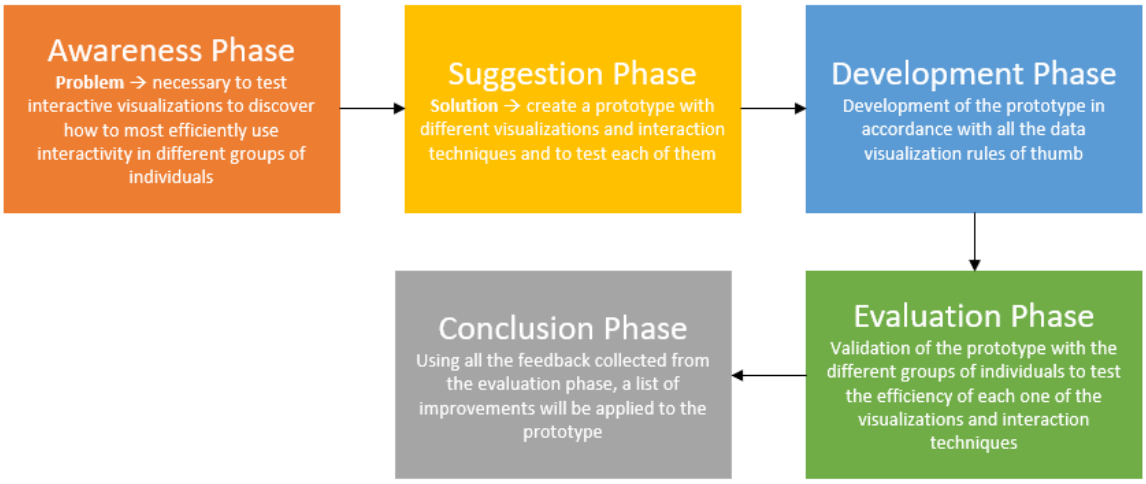


Figure 1 - Methodology Flowchart

The first stage is the awareness phase where it was necessary to understand what the underlying problem was. In this case, the problem was the fact that it was necessary to test interactive visualizations to discover how to most efficiently use interactivity in different groups of individuals. Additionally, and as mentioned previously, the motivation for the prototype topic was the fact that the Portuguese economy had a recent Financial and Assistance Programme motivated by external imbalances, which could be a very interesting subject to analyse through an interactive data visualization tool.

The second phase is the suggestion, and for this phase, the suggestion to solve the problem in question was to create a prototype with different visualizations and interaction techniques and to test each of them to see each which were the most effective. The third phase is the development, which consisted of the development of the prototype in accordance with the data visualization rules of thumb (Munzner, 2015). The next stage is the evaluation, which consisted of the validation of the prototype with the different groups of individuals to test the

efficiency of each one of the visualizations and interaction techniques applied in the prototype, in each group.

The last phase is the conclusion, where using all the feedback collected from the evaluation phase, a list of improvements will be applied to the prototype. The last three phases are going to be more extensively explained and discussed in the following sections.

Finally, to fully make sure that this prototype was developed into an efficient manner, the validation framework designed by Munzner (2015) was also applied to help in the design of the prototype. As mentioned in previous sections this framework contains four levels: Domain situation, Data/Task abstraction, Visualization encoding/Interaction idiom, and Algorithm.

In the top-level, it was analysed who were the main target users, which in this case were the experts and the novice's group, and which was the main area of interest of this prototype, which in this case was the balance of payments.

In the next level, all the data needed to build the prototype was collected and in the third level, all the visualization and interaction idioms were chosen for each one of the visualizations in the prototype, while complying with the data visualization rules of thumb. Finally, for the bottom level, the code developed to build this prototype was created in order to efficiently display all the visualizations.

3.1. DATA

As for the data source used to create the prototype, the dataset used was from *Banco de Portugal*. This dataset is public, and available at the *Banco de Portugal* website (Banco de Portugal, 2019). Even though the website of *Banco de Portugal* possessed data on several other different topics, considering that the motivation for the prototype was to analyse the Portuguese economy and its external imbalances, the dataset used only had data regarding the balance of payments (Banco de Portugal, 2019). This decision was made to prioritize the most important information in the prototype, which in this case is a Shiny application, in order to apply one of the suggestions in the work conducted by López *et al.* (2018).

A comparison of this website and the prototype was not made considering that this website was only used to retrieve the dataset used to build the prototype. Moreover, the fact that this website did not possessed any interactive visualizations was also another reason not

to compare the prototype with it, considering that the main goal of this website was to provide several chronological series of important statistics and data regarding the Portuguese economy and the goal of the prototype was to display data about a specific information of the Portuguese economy, in this case, the balance of payments, through the use of several interactive visualizations.

The dataset was pre-processed and then used to create several interactive visualizations that compose the prototype. It was also used an additional dataset from Eurostat. This dataset was used to create an international comparison of the Portuguese economy with all the countries in the European Union (Eurostat, 2018), using the current account balance in the percentage of GDP (gross domestic product), with a three-year average. This dataset was applied in the fifth visualization of the prototype "*By international comparison*".

3.2. TECHNOLOGY USED

As mentioned above, the prototype built in this study was developed using the R software and more specifically the Shiny package. R is a programming language and a free software environment for statistical computing and graphics and Shiny is a package that makes it possible to create interactive web base applications in R.

Regarding the R software in general, one reason to use this type of software is the fact that R is a software that exceeds other tools when it comes to flexibility (Jiang & Carter, 2018). Besides this, the fact that the software is free also proved to be a bigger advantage, in comparison to other tools that are used to create interactive visualizations, such as Tableau, since even though this tool is not as expensive as other tools it still has a cost (Bhardwaj & Baliyan, 2019). Another reason as to why this type of software was used in this project is the fact that R is a programming language that was created specifically for data analysis, on the contrary of the Phyton language that is a general-purpose programming language and not as specific as R. Although both could have been used, R also possess a smaller learning curved compared to Python and could be a better choice for beginners in both languages, since it can require less code than Python in some functions. Also, considering this project will not be used for production, and only in a standalone server, even though Python is faster than R, for the goal of this study, the speed that R achieves is suitable for the requirements of this project. Another interesting fact is that the R ecosystem offers a wide range of packages that can be used in data visualization.

The fact that R offers the possibility to use the Shiny package was also one of the key reasons to use this type of software. One of the motivations to use the Shiny package was the possibility to easily integrate interactivity in the applications created, since it provides several interactive features, which was one of the requirements of this project. This package can also be used together with other packages, which also provides an extensive variety of choices to use when creating a visualization. Besides that, the fact that it makes it possible to host these applications online was also a benefit since it would be feasible, if necessary, to test the prototype online instead of just testing it in an in-person experiment (Ellis & Merdian, 2015).

3.3. DEVELOPMENT OF THE PROTOTYPE

The prototype consists of a foreword view and five different interactive visualizations all regarding the topic of Balance of Payments. As mentioned before, all five visualizations were revised to fully achieve the requirements of this project and ensure that the data visualization rules of thumb were being applied (Munzner, 2015). For example, regarding the rules of thumb, no 3D was used, avoiding perspective distortion; 2D was used given that lists were not a suitable option for the different messages each visualization wanted to convene; a sidebar layout was used, with the filters on the left and on the right the main visualizations, in order to follow the rule "overview first, zoom and filter then details on demand", and the speed of the prototype was also taken into consideration since responsiveness was also an important factor.

Although all the five visualizations are about the topic of Balance of Payments, each one of them covers different sections of the topic so that the user interacting with the prototype doesn't have to go back and forth the views to understand a piece of specific information. This was done to prevent from causing irritation or annoyance in the users, and because combating oscillation is one of the suggestions in the work done by Blascheck, *et al*. (2019).

In the next sections, the explanation of the content in each view of the prototype, the storytelling, all the types of visualizations, interaction choices and data filtering, channels, marks and technical aspects of the prototype will be explained.

3.3.1. Content of each view of the prototype

The prototype starts with a foreword view where there is a video about the topic of this prototype that serves as an introduction.

Moving for the first view (*By major items)*, in this view it is explained and analysed each of the different items that compose the current and capital account of Portugal and the evolution of each item through the years, as well as the balance values of the current and capital account.

The subject of the analysis of the second view *(By geographical counterpart)* was the geographical breakdown of each one of the selected items per year. In this view the items were goods and services. The map represents, for the item/year selected, the relative value of each country in the world context, where the relative position of the counterpart economy was defined in terms of quartiles.

The main focus of the third view *(By monthly periodicity)*, was to analyse the patterns of each selected item for each month and year combination. The items available in this view were goods, services and the current and capital account.

In the case of the fourth view *(By type of services)* the goal of the analysis was to observe the evolution of the exports/imports of the different types of services, by analysing two different combinations of services, all types of services and all services except travel and transport.

Finally, in the fifth and last view *(By international comparison)* the analysis was focused on an international comparison of Portugal with different combinations of European countries. In this case the metric used to perform this international comparison was the current account balance.

It is also worth mentioning, the fact that the measure of unit used in all of the views was Euro Millions, with the exception of the fifth view, where the current account balance was measured using the percentage of the GDP using a three-year average.

3.3.2. Storytelling

The type of storytelling used in this interactive visualization was the interactive slideshow (Segel & Heer, 2010), considering that the user has a relevant level of interactivity

but there is still some messaging in the views. This way the user can use the information to further explore parts of each visualization.

Regarding the visual design, the main guidelines in this section were covered by using consistent visual frameworks, using established shots such as the messaging to situate the user and use pop-out features in the text, using color to reference information in the visualizations.

As for the messaging, headlines and captions in each view were used to facilitate the understanding of the visualization and with the help of the video in the foreword view it was possible to create engagement with the user at the start, giving a teaser of what the user should expect in the following views. The video also makes data relatable and puts numbers and facts into context with the use of questions and metaphors.

There was also a focus in trying to tell a story while showing the information, in order to support transitions, which is one of the suggestions in the work done by Blascheck, *et al.* (2019), since it will help the comprehension of the users and most likely increase the engagement.

Finally, regarding interactivity, awareness was given to not obscure data and to present the data in a clear form. The (adequate number of) interactive features are explicit, react to the users through the help of widgets and tooltips, and there are also suggested actions for the user in the text, which can be seen as SI cues that use feedforward (Boy, Eveillard, Detienne, & Fekete, 2016).

3.3.3. Types of visualizations

One of the suggestions when creating a Shiny application is to use different visualization methods/idioms in it (López et al., 2018), and considering that the prototype was built using the Shiny package, five different types of visualization idioms were applied in the prototype.

In the first visualization *(By major items)* it was chosen a stacked bar chart to encode the data, since that in the sample of the dataset used for this view, there is one quantitative attribute (current and capital account balance values) and two categorical key attributes (years and items that compose the current and capital account - goods, services, primary income, secondary income, and capital account). Also, one of the reasons to use this type of visualization was that it makes it possible to easily visualize trends (Munzner, 2015), which is

useful since one of the possible targets of this analysis could be to see the differences in each year/decade and structural features of Portuguese external imbalances. There is also a line chart encoded on top of the stacked bar chart, which encodes the balance values of the current and capital account throughout the years.

In the second visualization *(By geographical counterpart)*, considering that the dataset used had a geographical breakdown, it was possible to create a spatial analysis. In this case, it was used a choropleth map, that has one quantitative attribute - goods and services regarding values of exports, imports, and balance values. This type of visualization was chosen since one of the possible goals of this analysis could be to identify spatial clusters in the data, therefore which regions have higher or lower values of the selected item, and with a choropleth map is easier to achieve this goal considering that is one of their visual tasks (Zhang & Maciejewski, 2017).

In the third visualization *(By monthly periodicity)*, it was chosen to apply a heatmap visualization, to compare the differences of the values of the item selected (goods, services regarding values of exports, imports, and balance values and the current and capital account regarding values of debit, credit, and balance values) in the different months/years. It is possible to use this type of visualization since the data used had one quantitative attribute (the values of the item selected) and two categorical attributes, the months and the years. This type of heatmap does not order the data based on the clusters of the two categorical attributes since the goal of this analysis should be a time analysis comparing the differences in each year by month. This was also another reason to use this type of visualization since heatmaps are an effective way to represent time-varying analysis (Kumatani, Itoh, Motohashi, Umezu, & Takatsuka, 2016).

In the fourth visualization *(By type of services)*, it was chosen to do a scatterplot chart since the dataset had two quantitative value attributes, the value of exports and imports for each type of service. One of the motivations to use this visualization is that it makes it possible to find trends and outliers easily (Munzner, 2015) and this could be extremely useful since it may provide, using the example of this visualization, with the type of service that has a higher impact in the total value of the services.

In the fifth and final visualization *(By international comparison)*, the type of idiom used was a bar chart, considering that the data used had one quantitative value attribute, the

current account balance, that was measured using the percentage of the GDP using a three-year average and one categorical value attribute, the countries. This bar chart was also ordered from the country with the highest value of current account balance to the country with the lowest value, and this order was applied in every year displayed in the visualization. One of the motivations of using this visualization is that it provides an easier way of comparing values (Munzner, 2015), which could be very useful since one of the possible goals of this analysis could be to compare the values of Portugal with the rest of the different combinations of countries.

3.3.4. Interaction choices and data filtering

López *et al.* (2018) mentioned that one of the suggestions to apply when creating a Shiny application is to use different interactive methods, and as mentioned previously, considering that this prototype was built using the Shiny package, different interactive methods were applied in each of the visualizations that compose the prototype.

Each of the five visualizations could be decomposed into two components: the sidebar and the main panel. Additionally, for the prototype itself, it was chosen to apply a NavBar layout, where each of the visualizations is a tab in the application, to facilitate the interaction of the users with it and to benefit the storytelling of the visualization.

For each of the visualizations, the sidebar layout was chosen since it eases the analysis for the users, using graphic design guidelines such as left to right. This type of guidelines was taken into consideration in the organization of the widgets and all the visualizations, to facilitate the user's comprehension (Blascheck, et al., 2019). This was also done since leveraging spatial organization was one of the suggestions in the work of Blascheck, *et al.* (2019).

The users had three different widgets to choose from: slider widget, radio button widget, and select box widget. In some visualizations the user has a combination of two widgets - for example, in the fourth visualization *(By international comparison)* users can choose to use the slider widget and the radio buttons widget, whereas in other visualizations users can only have one widget - for example in the third visualization *(By monthly periodicity)* the user only has the select box widget.

The slider widget was used to help users perform a time analysis on the visualizations and in almost all of them it provides the users with a play button, so it is possible to show an animation of the changes in the data throughout the years. The play button also serves the purpose to invite the users to interact with the visualization, since inviting interaction is one of the suggestions in the work done by Blascheck, *et al.* (2019). In the animated sliders, it is also possible to filter by year, considering the users can stop the animation at any time and see the differences between each year. The only exception is the first visualization *(By major items)*, where the slider is used to select the range of years that the users want to see in the stacked bar chart, making it possible to also filter by year.

The select box widget was used to provide the users with more options for variables to display in the visualizations. In the second visualization *(By geographical counterpart)*, using the select box, the users can filter the data by choosing which variable they want to see displayed in the choropleth map. In this visualization, the users can select the item they want to choose, goods or services regarding either exports, imports or balance values (the difference between exports and imports). The same happens in the third visualization *(By monthly periodicity)*, except the variable selected will change the values in the heatmap and the users can also select the same variables as the second visualization, plus the current and capital account variables regarding either debit, credit or balance values.

The radio buttons widget was used to provide the users with different combinations of variables they may want to display in the visualizations. In the fourth visualization *(By type of services)*, users may choose if they want to see all the services in one graph or all the services excluding the travel and transport services that have a higher impact on the visualization. In the fifth visualization *(By international comparison)*, using the current account balance as a measure, users can compare the Portuguese economy with different combinations of countries, for example, all the European countries, Italy, Greece and Spain or countries that have entered the European Union after 2004.

All these widgets, when selected, make alterations in the visualizations, in order to allow the users to perform different types of analysis, either a time analysis or a combination of different variables. The widgets used SI cues applied to an external object, in this case, the widget itself, since the users have to interact with the widget in order to see changes in the visualization but to use the widgets they don’t have to use the actual graphs (Boy, Eveillard, Detienne, & Fekete, 2016).

In the main panel, the interaction choice the users had were the tooltips, which display visualization values. In the first visualization, the tooltips make it possible to see the specific values of each item of the current and capital account for any given year by moving the mouse over each bar and the balance values of the current and capital account over the years by moving the mouse over the line. In the third visualization, the tooltips were used to observe the specific values of each month and year combination for the item selected, by moving the mouse over each square of the heatmap. On the case of the fourth visualization, the use of the tooltips makes it possible to see more specific information about each service, for example, the value of exports and imports, the value of the weight (value of exports of that service divided by the total value of exports of all the services for that given year) and the name of the service, by moving the mouse over each bubble in the graph. Finally, in the fifth visualization, the tooltips are used to see the specific values of the current account balance for each country in any given year, by moving the mouse over each bar of the visualization. The tooltips used SI cues applied to the object of interest, in this case, the actual graphs, since to interact with the tooltips the users have to click on the graph itself (Boy, Eveillard, Detienne, & Fekete, 2016).

The only visualization that does not have this feature is the second (*By geographical counterpart*) considering that the R package used, *rworldmap*, does not provide this type of interactivity. Although tooltips are not possible to use in this visualization the trade-off was that this package made it possible to use the slider widget and perform a time analysis.

Finally, in the main panel of each visualization, there is always a small text to give a little introduction to the visualization with the purpose of facilitating the comprehension of the user and inviting interaction. In the fourth visualization besides the introduction text, there is also a more extensive description of each service, under the visualization, to satisfy the curiosity users may have. Both of these texts were added to the visualization considering that scaffolding complex interactions is one of the suggestions in the work done by Blascheck, *et al.* (2019).

3.3.5. Channels

In the stacked bar chart applied in the first visualization *(By major items)*, the magnitude channel used was the length to encode the values of the current and capital account, divided into their different components. The reason to use this type of channel is that the type of attribute used is an ordered attribute and this type of magnitude channel has a high

ranking in terms of effectiveness since it ranks in third place (Munzner, 2015). The identity channels used in this visualization were the spatial region, to show the values of the current and capital account ordered by the year's category and the hue, to categorize the different components of the current and capital account. These channels were chosen since they are the best identity channels to use in terms of effectiveness (Munzner, 2015).

Regarding the choropleth map in the second visualization *(By geographical counterpart)*, the magnitude channel used was the saturation, since the visualization used is a choropleth map that uses color to encode ordered attributes and this channel is directly correlated with the type of visualization being applied. In this case, the saturation encodes the value of the item selected for the different countries presented, where the darker colours are the higher values and lighter colours the lower values. Although the map creates four distinct categories for each different color, since the attribute is ordered, the channel used is saturation and not hue. The identity channel used was the spatial region, to encode the different countries. This channel was used considering there was a categorical attribute, the countries, and since the dataset has geographical data and the type of visualization is a choropleth map, it is possible to use the spatial region channel and that is the identity channel that ranks the highest in terms of effectiveness (Munzner, 2015).

In the heatmap presented in the third visualization *(By monthly periodicity)*, the magnitude channel used was the saturation, considering that this type of visualization is a heatmap and uses color to encode information. Another reason to also use this type of channel is that this channel is also directly correlated with the type of visualization being applied. In this case, the saturation encodes the value of the item selected for the different months and years present, where the red colours corresponded to lower values and the green colours corresponded to higher values. The identity channel used was the spatial position, to show the values of the item selected ordered by the years and months categories. This channel was used since it is the highest ranked identity channel to use in terms of effectiveness (Munzner, 2015).

Concerning the scatterplot chart in the fourth visualization *(By type of services)*, the magnitude channel used was the position on a common scale, considering that the attributes used were ordered, since the position of each type of services is calculated given the quantitative values of the imports and exports and their value on the scale. Another magnitude channel used was the area, considering that the bigger the value of exports, the bigger the size

of the point was. The position on the common scale was used to encode the position of each dot since that is the most important information on the visualization and this channel is the best magnitude channel in terms of effectiveness (Munzner, 2015). The area was used to encode the value of exports, since it is a high-ranking magnitude channel in terms of effectiveness and was the most reasonable channel to apply considering what was being encoded and the type of visualization used. The identity channel used for this visualization was the hue, to categorize the different types of services and to highlight the Travel and Transport service. This channel was used considering it was not possible to use spatial region and hue is the most effective identity channel after that one (Munzner, 2015).

Regarding the bar chart used in the fifth and final visualization *(By international comparison)*, the magnitude channel used in this visualization was the position on a common scale, to show the values of the current account balance using the percentage of GDP, a three-year average, ordered by the country category. This channel was used since it is the highest-ranking magnitude channel in terms of effectiveness. Another magnitude channel used was the length to encode the values of current account balance using the percentage of GDP, a three-year average, for each country. The reason to use this type of channel is that the type of attribute used is an ordered attribute and this type of channel has a high ranking in terms of effectiveness since it ranks in third place compared to other magnitude channels (Munzner, 2015). The identity channel used was the hue, to categorize the countries and highlight Portugal. It was chosen to highlight Portugal since these are interactive visualizations about the Balance of Payments regarding Portugal but also because the users participating in this experiment will most likely be Portuguese and this will be an entry point for them since it is a more familiar country and it will likely increase the engagement with the visualization. This was also done since providing entry points it is one of the suggestions in the work done by Blascheck, *et al.* (2019). This channel was chosen since it is one of the best identity channels to use in terms of effectiveness, ranking in second place (Munzner, 2015).

3.3.6. Marks

Regarding the stacked bar chart used in the first visualization *(By major items)*, the marks used were points and lines. The lines represent two things, they can represent each bar, encoding information relative to the value of the current and capital account divided by the different components and they can also encode the information of the balance values of the current and capital account. The points represent each year in the line.

Concerning the choropleth map used in the second visualization *(By geographical counterpart)*, the only marks used were the areas to encode each region.

In the heatmap used in the third visualization *(By monthly periodicity)*, the only marks used were the areas to encode each value of the item selected by month and year.

In the scatterplot chart used in the fourth visualization *(By type of services)*, the marks used were points and lines. The points encoded the position of each service considering the value of the imports and exports and the size considering the weight for each type of service. The line encodes a bisector line that makes it possible to see if each type of service has bigger values for imports or for exports, or in other words if it contributes positively or negatively for the current and capital account. In this case if a bubble (service) is below the line, that service has for that given year, a value of exports higher than imports.

Finally, in the bar chart used in the fifth view and final view *(By international comparison)*, the only marks used were the lines to represent each bar. Each bar encodes information relative to the value of current account balance using the percentage of GDP, a three-year average, for each country.

3.3.7. Technical aspects

To build the Shiny app it was necessary to first build a pre-process R script that prepared the data to be used in all the visualizations. Then an R script was used where a user interface and a server were created.

The user interface was where all the layout was made and, in the server, all the operations and computations were made. The Shiny package was used in all the views to provide interactive visualizations that could be published on the web as well as to provide with the user interface-server architecture.

Besides the Shiny package, the remaining packages used in each visualization can be seen in **Table 3**.

Table 3 - R packages used in each one of the interactive visualizations

Visualizations / Packages	1º visualization	2º visualization	3º visualization	4º visualization	5º visualization
Plotly	**X**		**X**	**X**	**X**
Rworldmap		**X**			
RColorBrewer		**X**	**X**		
Ggplot2				**X**	

The package *plotly* was used in the first visualization *(By major items)* to create the interactive stacked bar chart. This package was also used in the third visualization *(By monthly periodicity)* to create the interactive heatmap and in the fourth visualization *(By type of services)* to create the interactive scatterplot together with the package *ggplot2*. Finally, it was also used in the fifth visualization *(By international comparison),* to create the interactive bar chart.

The package *RColorBrewer* was used to create a color pallet for the visualizations in the second visualization *(By geographical counterpart)* and in the third visualization *(By monthly periodicity).*

The package *rworldmap* was used in the second visualization *(By geographical counterpart),* to create a choropleth world map.

Lastly, HTML was used to write all the text present in the visualizations.

In the next chapter it will be possible to observe the visual representation of each one of tabs that forms the prototype.

3.4. EVALUATION OF THE PROTOTYPE

Succeeding the development of the prototype, the evaluation phase was then implemented. The research design applied in this phase was a causal-comparative research design since the research strategy used was an experiment that tried to find connections between independent variables, in this case, each of the different types of visualizations and interaction methods applied, and other dependent variables, using two different groups of individuals (Schenker & Rumrill, 2004).

The experiment was in most cases an in-person experience and was conducted with a sample of individuals, from groups of experts and novices. The group of experts are

individuals that have knowledge about the topic in the interactive visualizations, in this case, the balance of payments. This type of individuals was recruited by contacting individuals that work in *Banco de Portugal*, professors and individuals that study or work in the economy field or other similar areas. The group of novices are individuals that have no prior knowledge about the topic of the balance of payments. This group was recruited by contacting students from universities and individuals that study or worked in different areas, that could participate in the study. Both groups were tested to realize if the interactive visualizations were understandable and efficient for the different types of individuals (Blascheck, et al., 2019). It was also ensured that both groups had individuals from different ages from 18 to over 55 and individuals from different genders.

The conceptual model that was used in the experiment was a graphic model and can be seen in **Figure 2.**

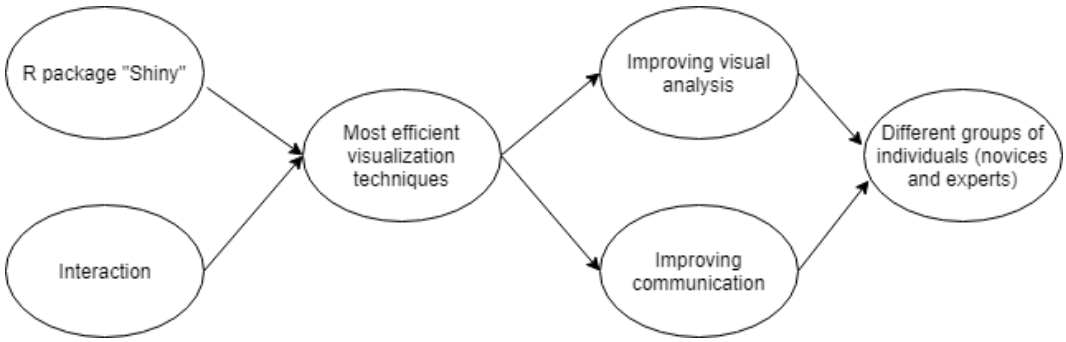


Figure 2 - Graphical Conceptual Model

To fully make sure that the validation of the prototype was implemented correctly, considering that the four-levels of design of the framework created by Munzner (2015) were applied when designing the prototype, the four-levels of validation were also applied in this stage. To do this, considering the top level of this framework, the Domain situation, it was necessary to understand if the correct target users were being applied, and considering that two different types of individuals were needed, experts and novices, the only requirement that needed to be validated was making sure that the individuals from the groups of experts had knowledge on the topic of the visualization, by requiring work experience or a degree in the field of economy or similar areas. Regarding the group of novices, the only requirement was that this group should not have any previous knowledge on the topic of the visualization.

In the next level, Data/Task abstraction, it was necessary to understand if the requirement concerning the data used in this prototype was being fulfilled, which was already

validated considering that the data used was from the *Banco de Portugal* website, which is a reliable and updated source of information that contains data regarding the main interest of the prototype.

The third level, Visual encoding/Interaction idiom, contained the most extensive part of the validation of the prototype. It was at this level, that the prototype was evaluated with the two different types of individuals, performing the experiment of this study and using quantitative and qualitative measures to perform the evaluation of the prototype (Lu, et al., 2018).

Finally, in the fourth level, Algorithm, the requirement to validate this stage was to ensure that the code used to develop the prototype was deployed effectively and displayed all the components of the prototype, and that if the experiment was online instead of in-person, the website created to deploy the prototype was working functionally.

Concerning the experiment of this study, the goal was to test which one of the visualizations and interactions choices in the prototype were the most effective and the differences between how the different groups of individuals interacted with the prototype. In the case of an in-person experiment, the participants were observed during all the time it took to complete the experiment. To start the experiment, the sample group, which consisted of a minimum of thirty participants, fifteen for each group, had access to the prototype and was able to interact with each one of the interactive visualizations. After a small period for testing had passed, the participants had to follow a specific set of tasks and record their answers in a questionnaire (Géryk, 2015), named "Experiment tasks" (**Annex A**) to fully understand if they were able to use the prototype to the maximum of its potential.

Considering that, as mentioned previously, most experiments were an in-person experience, almost all the participants were observed while performing the experiment. This forced the participants to have an active attitude when interacting with the prototype, fulfilling one of the suggestions to apply when creating a Shiny application (López et al., 2018).

In this part of the experiment, using the "Experiment tasks" questionnaire it was possible to collect the quantitative measures (Lu, et al., 2018), such as the accuracy (Zhu, 2007), by comparing the correct answers with the answers of each participant and the completion time (Zhu, 2007), by measuring the time it took each participant to finish the tasks

given in each of the interactive visualizations and the total time it took to finish this part of the experiment.

In the next stage of the experiment, the participants had to answer a 5-point Likert scale questionnaire named “Qualitative questions” (**Annex B**), related to their opinion of the prototype and were able to suggest improvements on each one of the interactive visualizations and its components. In this part of the experiment, using the “Qualitative questions” questionnaire, each section of this questionnaire had questions that measured the level of usefulness, effectiveness, complexity (Zhu, 2007), and engagement (Lu, et al., 2018) that the participants classified each one of the visualizations, as well as a final section regarding the prototype as whole. The participants also classified the level of usefulness and complexity of the components of each visualization. This made it possible to collect each one of these qualitative measures (Lu, et al., 2018). In the final part of the “Qualitative questions” questionnaire, the participants were also asked to rank each one the visualizations that compose the prototype. All the measures collected in both questionnaires were analysed and considered in the improvement of the prototype.

In the special cases in which the participants could not attend the in-person experiments it was still feasible to test the prototype remotely since, as mentioned above, with Shiny it is possible to create web page applications (Ellis & Merdian, 2015), and because of this, it is possible to access the prototype through a link. In these cases, the participants only had to document the time they took to complete each of the questionnaires. In this type of online experiment, the only thing that was not included in the prototype was the video in the foreword, to increase the speed of the experiment.

An in-person experiment was preferred since it was possible to document the choices that the participants applied during the tasks phase, to better understand how each of the participants interacted with each of the interactive visualizations of the prototype. These experiment observations were also taken into consideration in the improvement of the prototype. It was also possible to measure the time each participant took to conclude the tasks on each interactive visualization instead of only measuring the time it took to complete the full questionnaires. It also increased the fidelity of the data, since the participants were observed during the experiment. Additionally, if the experiment was in-person the local mode of R was used, increasing the speed of the prototype when compared to the web mode. The

in-person experiment was also ideal since, in the case of any errors happening, these could be solved, and the participants would still be able to complete the experiment.

The dependent variable of this study was the efficiency in how the individuals from the sample used the different interactive visualizations of the prototype and how well they understood the message being transmitted by them. This dependent variable can be studied by analysing all the specific measures collected, for example, the quantitative measures, such as the accuracy in how the individuals used each one of the interactive visualizations and perform the tasks needed, as well as the completion time. Besides these quantitative measures, the qualitative measures also need to be analysed, such as the usefulness, the complexity, the engagement and the effectiveness of each visualization and its components (Géryk, 2015; Lu, et al., 2018; Zhu, 2007). This process was done by analysing both questionnaires applied in the experiment as well as the experiment observations documented in the in-person experiments.

In the analysis of the results, the mean values of each one of the measures (Géryk, 2015) obtained in both questionnaires and the experiment observations, were analysed to develop the conclusions of this study. An ANOVA single factor test was also conducted to test for statically significant differences between both groups.

The equipment and set up used in the in-person experiments were the computer HP Pavilion Gaming Notebook - 15-ak006np, with a 15 inches screen. The in-person experiments were conducted using the local mode of the R software with an inner window size of 1536 x 754, in a room isolated from outside distraction.

To conclude, all the information collected was used to improve the prototype and to reach conclusions about what are the best visualization and interaction techniques to improve visual analysis and communication in different groups of individuals and to find differences between how these different groups perceive interactivity, using the tool mentioned above.

4. RESULTS

In this chapter the results of the experiments done in this project will be documented. The results consisted of the prototype created and all the measures collected in the experiments: the quantitative measures, such as the accuracy and the completion time, and the qualitative measures, such as the usefulness, the complexity, the engagement and the effectiveness. Finally, it will also be in this section that the experiments observations collected in each experiment will be analysed in order to develop the improvements for the prototype.

4.1. VISUAL REPRESENTATION OF THE PROTOTYPE

The first result of this study was the prototype, that allowed the implementation of the experiments of this project and the results that therefore followed. Thus, it is important, to first show the visual representation of the prototype, before starting to analyse the results of the experiment.

The foreword page of the prototype can be seen in **Figure 3**.

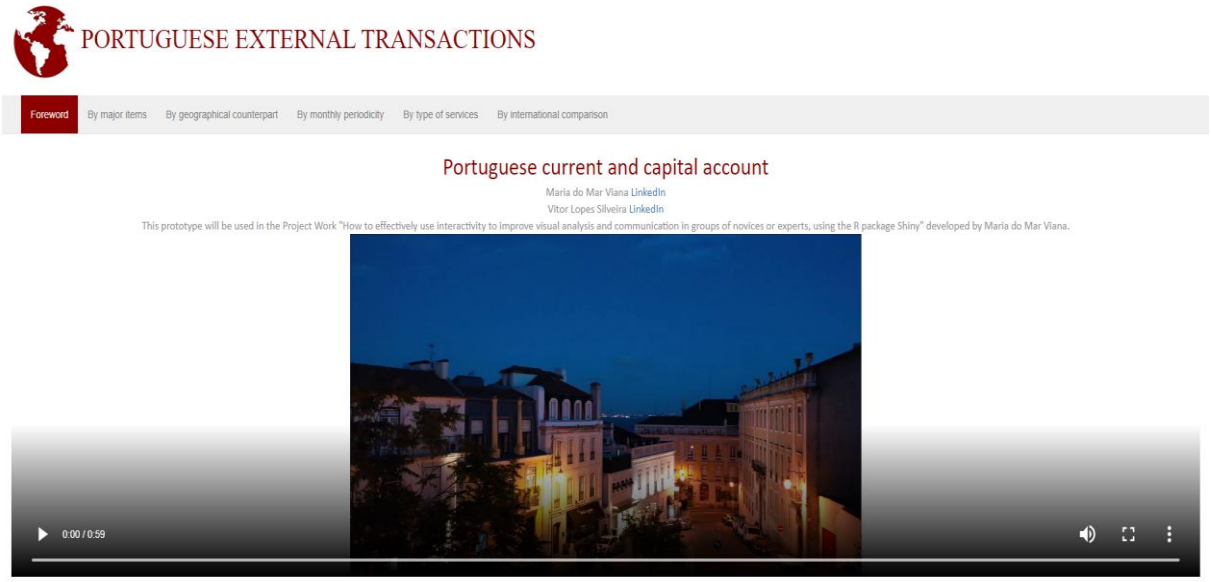


Figure 3 - Prototype "Foreword" tab

The "*By major items*" tab, which corresponds to the first visualization, where the idiom is a stacked bar chart can be seen in **Figure 4**.

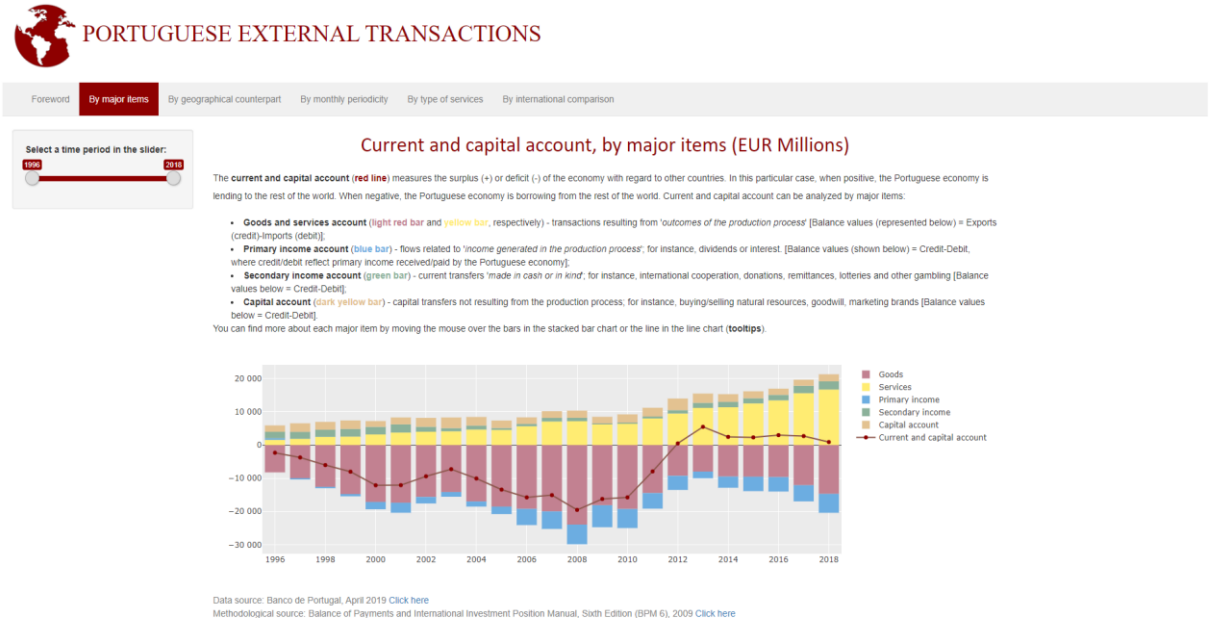


Figure 4 - Prototype "By major items" tab

The "*By geographical counterpart*" tab, which corresponds to the second visualization, where the idiom is a choropleth map can be seen in **Figure 5**.

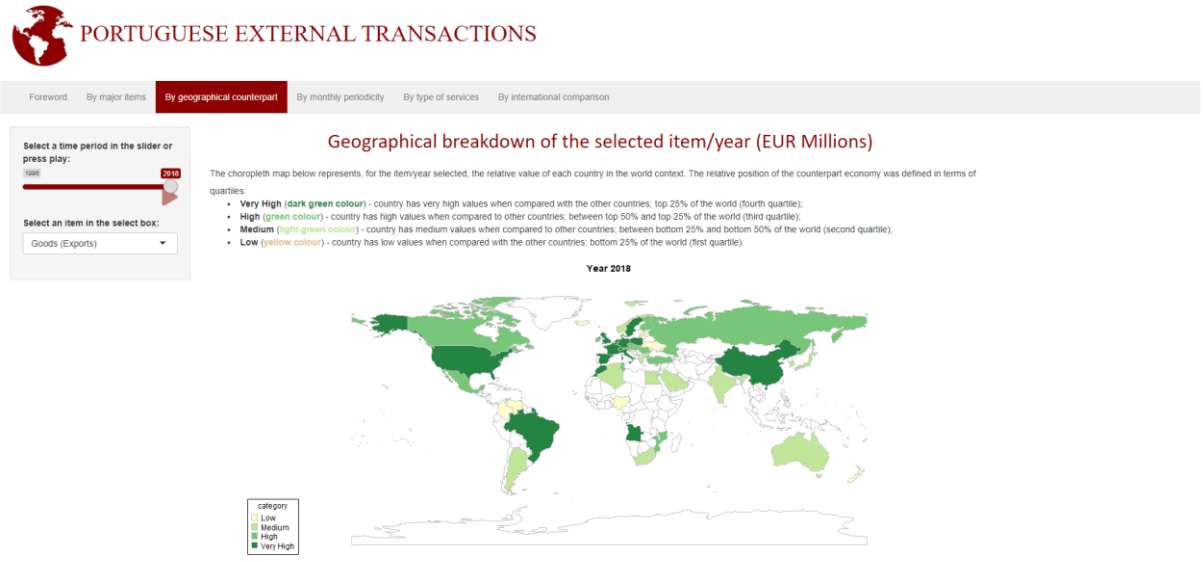


Figure 5 - Prototype "By geographical counterpart" tab

The "*By monthly periodicity*" tab, which corresponds to the third visualization, where the idiom is a heatmap can be seen in **Figure 6.**

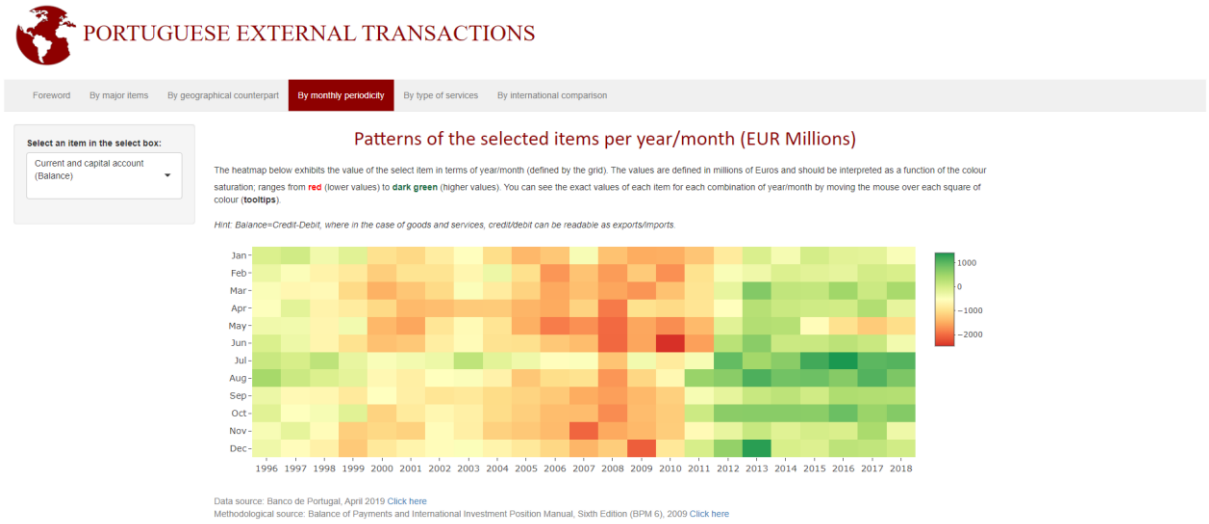


Figure 6 - Prototype "By monthly periodicity" tab

The "*By types of services*" tab, which corresponds to the fourth visualization, where the idiom is a scatterplot chart can be seen in **Figure 7.**

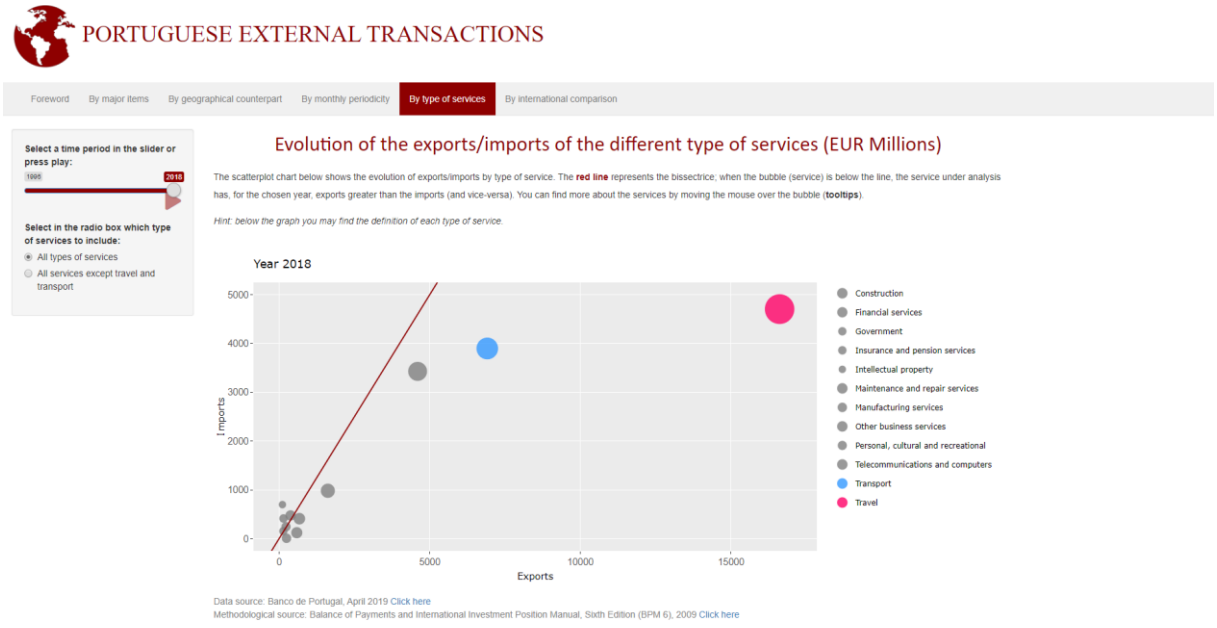


Figure 7 - Prototype "By types of services" tab

Finally, the "*By international comparison*" tab, which corresponds to the fifth visualization, where the idiom is a bar chart can be seen in **Figure 8.**

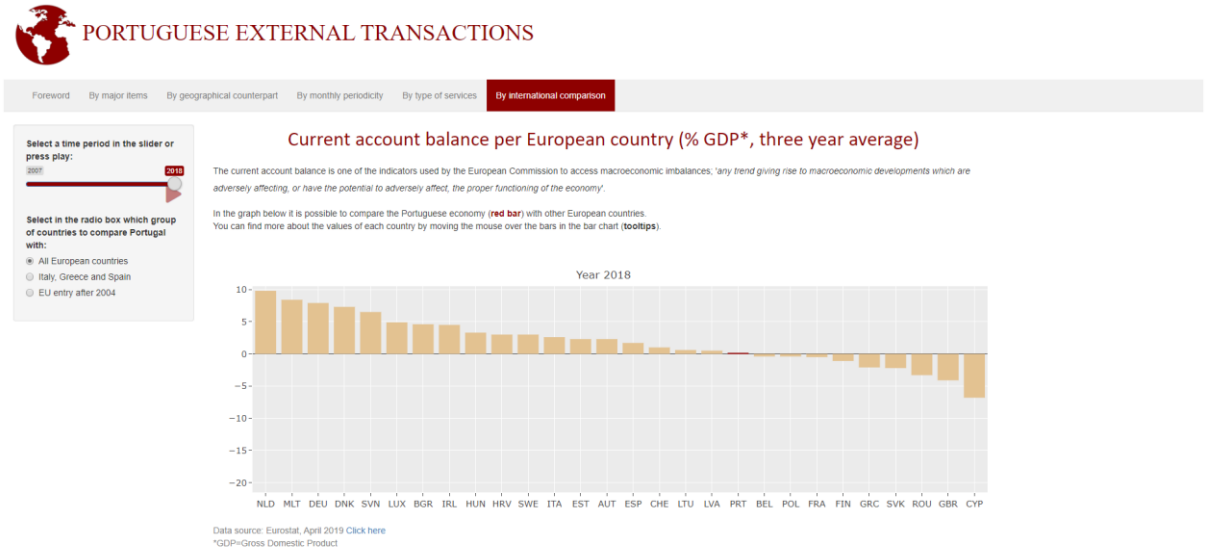


Figure 8 - Prototype "By international comparison"

4.2. EXPERIMENT RESULTS

To start the analysis of the experiments results, it is first worth mentioning that, as mentioned previously, 30 individuals participated in this experiment, 15 in the group of experts and 15 in the group of novices, making this sample statistically significant considering it used more participants than the studies mentioned in the literature review, since for example the study conducted by Blascheck, et al (2019) used 24 participants, 8 in each group and in the works conducted by Oghbaie, Pennock, & Rouse (2016) and by Rouse, Pennock, Oghbaie, & Liu (2017) only 10 participants were used in total.

Additionally, for all of this measures the mean results of all the participants in both groups, experts and novices, were calculated, in order to facilitate the analyses and the comparison of the performance of both groups. The mean was used to compare the results considering that when using the mean all the values in the data set are included in the calculations and an adjustment in any of the data will influence the estimation of the mean. Finally, as mentioned previously, an ANOVA single factor test will be used to test for statically significant differences between both groups.

4.2.1. Quantitative measures

The first measures that were documented and analysed were the quantitative measures, which as mentioned previously, were the accuracy and the completion time. Regarding the quantitative measures the first that was analysed was the accuracy. This measure was obtained by dividing the number of correct answers that each participant gave in the questionnaire "Experiment tasks" by the total number of questions of this questionnaire in order to obtain a percentage, considering that this measure was analysed using a scale of 0-100%. It was created a matrix that can be seen in **Table 4**, to divide the results into four intervals, Very low, Low, Moderate and High, in order to classify the results of the accuracy measure.

Table 4 - Matrix of classification of the accuracy measure

INTERVAL OF VALUES	CLASSIFICATION OF THE MEASURE
[0-25%]	Very low accuracy
[26%-50%]	Low accuracy
[51%-75%]	Moderate accuracy
[76%-100%]	High accuracy

The first results that were obtained were that in average, the accuracy of the experts is higher than the accuracy of the novice's group. The experts group obtained a global average of 93% accuracy, whereas the novices group obtained 91% accuracy. The mean values of the accuracy of each one of the visualizations for both groups were also obtained and can be observed in **Figure 9**.

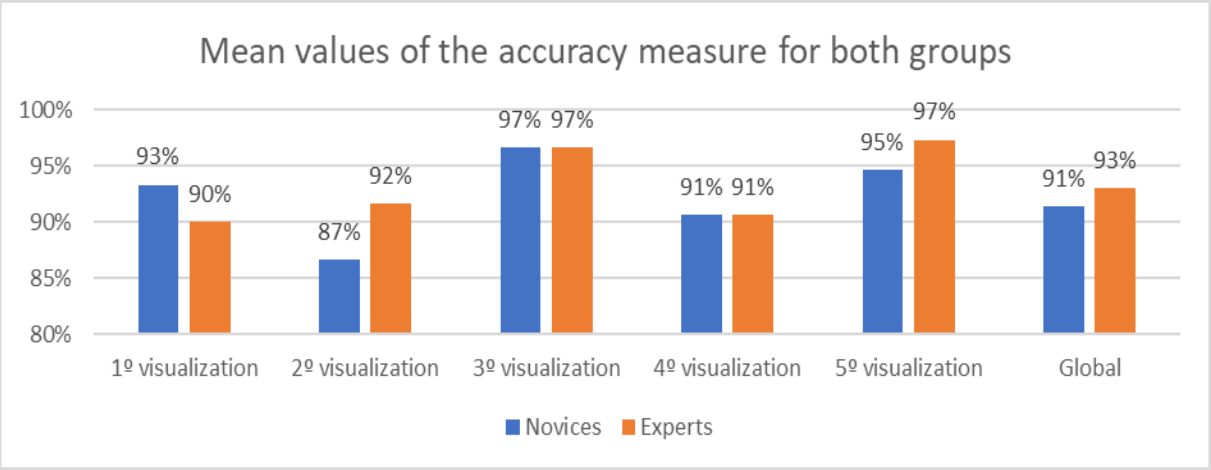


Figure 9 - Mean values of the accuracy measure for both groups

As can be seen in **Figure 9**, it is possible to observe that the experts group had in most of the visualizations of the prototype, a higher or equal average accuracy than the novice's

group. This observation has logic considering that the experts group has more knowledge on the subject of the visualization than the novice's group, and it would therefore be reasonable to think that they would have higher mean results.

Nevertheless, the novice's group average accuracy was higher than the expert's group in the case of the first visualization, which can be extremely interesting considering that in the experiment observations, the participants of the novice's group usually had difficulties with this visualization. Another interesting fact is that although they had lower results than the experts group in second and fifth visualization, they still had a higher average of accuracy in those visualizations and had the same accuracy than the experts group in the third and fourth visualization, where both of those averages of accuracy were in the High accuracy interval. In fact, in both of the groups, the average accuracy in all the visualizations were above 75%, which means that all the results were in the High accuracy interval. This may indicate, that although the participants of the novices group did not have any previous knowledge on the subject of the visualizations, they were able to answer the questions given to them and infer knowledge with just the information present on the visualizations, closing the gap between the experts group and them.

Also, regarding the accuracy, another calculation that was possible to take using the questionnaire "Experiment tasks", was the global average number of wrong questions, which as expected, the experts group had the lowest number of wrong questions considering they had a higher global average accuracy. The average number of wrong questions in the expert's group was 1.67, whereas in the novices group their average number of wrong questions was 1.87.

Another observation could be that, the only visualization where there was a more noticeable difference of accuracy between the groups was in the second visualization, where the novices group had an accuracy of 87% and the experts group an accuracy of 92%. While both of these results are still considered to be in the High interval, this may mean that the novices group has difficulties analysing choropleth maps, considering it was the visualization with their lowest average of accuracy and that was the idiom chosen for the second visualization. The lowest average of accuracy for the group of experts was in turn in the first visualization, which was a stacked bar chart, which can be interesting considering that this a very common idiom to use in economy, but nevertheless their average was still considered to be in the High interval.

Finally, as for the highest averages of accuracy, in the novice's group, the visualization with the highest value was the third one, where the idiom was an heatmap, and that could be interesting considering that this type of idiom can be difficult to analyse for individuals with no knowledge on the subject. This was also the visualization with the highest average of accuracy for the experts' group, as well as the fifth visualization, which was a bar chart, that also possessed the same average of accuracy of 97%, which was the highest value of average of accuracy for both groups.

The final quantitative measure being analysed was the completion time, which was obtained by measuring the time it took every participant to complete the questions for each visualization and the total time it took them to complete this stage of the experiment.

On the contrary of the accuracy measure, the group that had the lower global average, of completion time - which in this case the lower the more efficient - when performing the experiment was the novices group, with a global average of completion time of 21 minutes and 57 seconds and the experts group with a global average of 24 minutes and 1 second. The averages of completion time for each one the visualizations were also measured and can be seen in **Figure 10**.

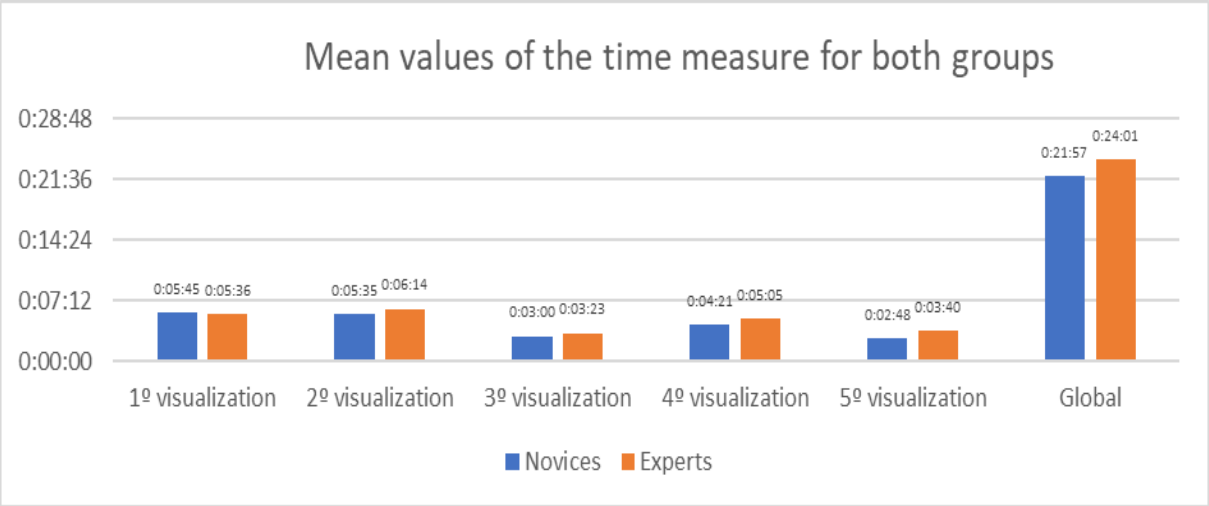


Figure 10 - Mean values of the time measure for both groups

As it is possible to observe in **Figure 10**, the novices group had lower average times for all the visualizations, except for the first visualization. These results are coherent with the results in the accuracy measure, since the first visualization was the only visualization where the novices group had a higher accuracy than the expert's group. This may be because they spent more time on it and tried to find more data and information to answer the questions

given to them. The same observation can also be the reason why the experts had higher or equal averages than the novice's group in other visualizations, since they spent more time on them, and tried to possibly discover more data in the visualizations than the novice's group.

Another interesting observation is, that the difference between the times of each group is never higher than 1 minute in any visualization. This may mean, that although the experts group took more time analysing the visualizations and that may have influenced their accuracy, the novices group still has similar times to them, and that may have influenced their accuracy as well, considering both of the groups have high averages of accuracy in all of the visualizations.

Lastly, it is also worth noticing that the visualization that had the lowest average of completion time for the group of experts was the third visualization, which is interesting considering that the third visualization idiom is an heatmap. Although for the group of novices the visualization with the lowest completion time was the fifth visualization where the idiom was a bar chart, the third visualization also had one of the lowest completion times and considering this visualization was also the one that had the highest average of accuracy for both groups, this visualization was likely the easiest one to analyse in all of the prototype. On the contrary, the visualization with the highest average of completion time was the second visualization for the expert's group, which may also mean that even groups with knowledge about a subject may find that it takes longer to understand a choropleth map. The second visualization was also the visualization that had the lowest average of accuracy for the novice's group, which means that both groups had difficulties on this visualization. On the novices group the visualization with the highest average of completion time was the first visualization, which is a stacked bar chart, which in turn was the visualization where the group of novices had their highest average of accuracy, perhaps because they took more time to analyse it.

Finally, to test for significant effects, an ANOVA single factor test was conducted regarding the quantitative measures. Regarding the accuracy measure, that can be seen in **Table 5,** it was not possible to find differences between the two groups that were statistically significant ($F(1,8) = 0.122605$, $p = 0.735266$). The same results were obtained regarding the completion time measure, considering no statistically significant differences were found ($F(1,8) = 0.358601$, $p = 0.565851$), as can be seen in **Table 6**. This means that although there seemed to exist differences between the quantitative measures in both groups these

differences are not statistically significant, meaning there are no differences in both groups regarding the accuracy and the completion time measures.

Table 5 - ANOVA single factor test for the accuracy measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	4.63	0.926	0.00148
Experts	5	4.67	0.934	0.00113

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.00016	1	0.00016	0.122605	0.735266	5.317655
Within groups	0.01044	8	0.001305			
Total	0.0106	9				

Table 6 - ANOVA single factor test for the completion time measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	0.014919	0.002984	9.27E-07
Experts	5	0.016644	0.003329	7.31E-07

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	2.97E-07	1	2.97E-07	0.358601	0.565851	5.317655
Within groups	6.63E-06	8	8.29E-07			
Total	6.93E-06	9				

The last observation that was possible to make using the questionnaire “Experiment tasks”, regarding the quantitative measures, was the analysis of which was the question in all of the prototype that had the highest number of wrong answers to it, and this question was question number four in the first visualization. The reason for this could be the fact that the first visualization was the visualization with the lowest accuracy for the experts group, but this question was also the question with the highest number of wrong answers for the novices

group, that had one of the highest average of accuracy in this visualization, meaning that may not have been the case. The most reasonable explanation is likely the fact that this question was the only one that had multiple correct answers, which may have come unnoticed for several participants, either in the group of novices or in the group of experts.

This is important considering that if the participants only selected one of the corrected answers in this question, the question was seen as wrong, considering that there was an option that included both of the corrected answers, and that was the correct one.

In **Table 7** it is also possible to see which question had the highest number of wrong answers for each visualization in each group.

Table 7 - Question with the highest number of wrong answers

Group / Visualizations	1º visualization	2º visualization	3º visualization	4º visualization	5º visualization
Novices	**4**	**9**	**13**	**17**	**24**
Experts	**5**	**10**	**14**	**18**	**22**

4.2.2. Qualitative measures

The next measures that were analysed were the qualitative measures, respectively, the usefulness, the complexity, the engagement and the effectiveness. These measures were compared using their average values for each visualization, same as in the quantitative measures. Besides analysing these four qualitative measures for each visualization, it was also analysed the measures of complexity and usefulness for each component of each visualization. At the end of the questionnaire "Qualitative questions" it was also asked to classify the usefulness, engagement and effectiveness of the prototype as whole, but considering these measures were already retrieved for each individual visualization, these measures were not taken into consideration and the global values of the prototype were calculated using the average values of each visualization.

All these measures were measured using a 5-point Likert scale and they were classified using the matrix that can be seen in **Table 8**, which divided the results into four intervals: Very low, Low, Moderate and High.

Table 8 - Matrix of classification of the qualitative measures

INTERVAL OF VALUES	CLASSIFICATION OF THE MEASURE
[1-1.99]	Very low
[2-2.99]	Low
[3-3.99]	Moderate
[4-5]	High

In the final part of the questionnaire "Qualitative questions" the participants were also asked to rank each visualization of the prototype, where "1" was the first place and therefore the visualization that was the most preferred, and "5" was the last place and the least preferred visualization. These values were also averaged in order to create the average rank for each visualization and see which was the most preferred and least preferred in each group, novices or experts.

The first qualitative measure that was analysed was the usefulness, which can be understood as the usefulness of the visualization or the prototype in answering the questions asked in the experiment and how useful was each visualization for the global analysis. The results obtained can be observed in **Figure 11**, and one of the first observations can be that all of the average values of the usefulness are in the fourth interval, meaning that for each visualization and in global, both groups classify the usefulness as High, although the novices group had a higher global average of usefulness than the experts group, with a value of 4.86 and the experts group of 4.75.

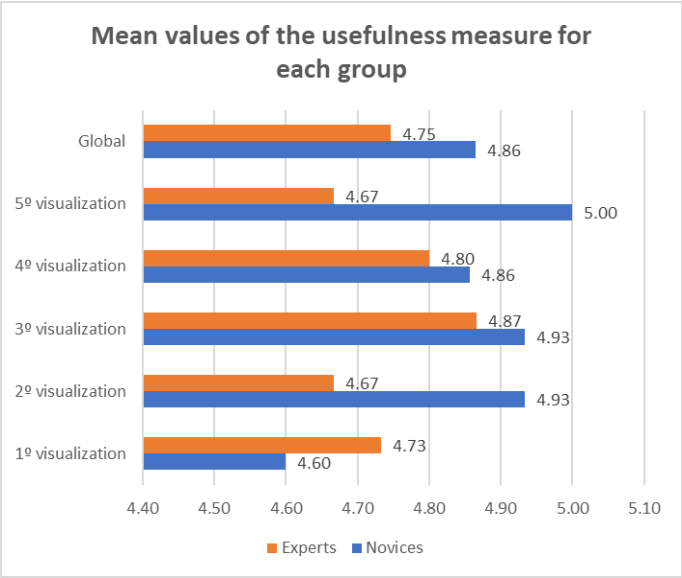


Figure 11 - Mean values of the usefulness measure for each group

Another interesting observation is that for most of the visualizations, the novices group had higher values of the average of the usefulness than the expert's group, except for the first visualization. This means that in most of the cases the novice's group thought that the visualizations were more useful to answer the questions asked than the expert's group. This can be due to the fact that the novices group had less knowledge than the experts group, and were therefore more reliant on the visualizations to answer the questions, on the opposite of the experts group, where for example, some participants in some questions already knew the answers without looking at the visualization, merely based on the previous knowledge they already had. It can also be because the expert's group may be more demanding than the novice's group, considering they already had previously knowledge on the subject.

It is also interesting to notice that the only exception, as previously mentioned was the first visualization, where the experts group had a higher average value of usefulness than the novice's group. This may be due to the fact that this was also the visualization where the novices group obtained the lowest average of usefulness, which is interesting considering that the idiom of this visualization was a stacked bar chart, and regarding the accuracy, the novices group scored a relatively high average of accuracy on this visualization. This means that although they had a higher value of accuracy, they didn't find it as useful as the other visualizations, perhaps due to the fact that this visualization is the visualization that has more text and that might make it less useful than others. It is also interesting to notice that the visualization that had the highest average of usefulness for the novices group was the fifth visualization, which makes sense since they had a higher average value of accuracy and the idiom of the fifth visualization is a bar chart, which could be simpler than other visualizations considering they usually are a more common type of idiom and the participants may have been more familiarized with it. Another interesting observation is although the fifth visualization had the highest average of usefulness, the second visualization still obtained a very high average of usefulness with a value of 4.93, which is interesting considering that this was the visualization where the novices group had the lowest value of accuracy, with a value of 87%. This can mean that although they didn't get as much correct answers in the second visualization, they still thought it was very useful, perhaps because they were not aware that they were answering incorrectly.

As for the group of experts, the visualization that had the highest average value of usefulness was the third visualization, where the idiom was a heatmap. This has logic,

considering this was the visualization where the group of experts had the highest value of accuracy. The visualizations where the experts group had the lowest average of usefulness were the second and the fifth visualization, which for the second visualization may be due to the fact that this was the visualization that had the highest completion time for the experts group and since they took longer analysing it, they may have found it less useful than other visualizations because of that fact. As for the fifth visualization, this was one of the visualizations that also had the highest accuracy for the groups of experts, meaning that accuracy may not always be correlated with the usefulness. One reason may be due to the fact that bar charts are usually and as mentioned previously, a commonly used idiom and on the experiment observations some participants mentioned that they thought the fifth visualization was to simple, so they may not have found it was as interesting as other visualizations and that therefore decreased its usefulness in the analysis. It is also possible to observe the relatively great difference in how both groups classified the usefulness of the fifth visualization, meaning that for the group of novices it may have had a higher usefulness due to the fact that they were more familiarized with it and for the experts group it had one of the lowest average of usefulness since they may have found it was too simple and not as useful as the other visualizations.

The next measure that was analysed was the efficacy, which can be correlated to the performance of the participants, or otherwise the accuracy. It is therefore linked to the level of helpfulness that the participants thought the visualization had in helping them select the correct answers. The results obtained can be seen in **Figure 12**, and the most relevant observation is that, as similar to the usefulness measure, all the values are above 4, which means they all ranked in the fourth interval of High efficacy. The global values of average efficacy were very similar for both groups, but the novices group had a slightly higher average, with a value of 4.84 and the expert's group of 4.81. This can be interesting considering that the experts group had a higher value of accuracy than the novice's group. This means the accuracy measure was the actual performance of groups and the efficacy measure, measures how well they thought their performance was in each visualization. This can be related with a phenomenon already previously mentioned, the fact that the novice's groups may have not been aware they were answering some questions wrong, and they thought they were correct.

As similar to the usefulness measure, the novice's groups tended to have higher averages of efficacy than the experts group on most visualizations, with again, the exception of the first visualization, being coherent with the results of the usefulness measure, since it would make sense that these two measures would be correlated. The first visualization also happened to be the visualization that also had the lowest average of efficacy for the novice's group, which is interesting considering that this group had, as previously mentioned, a relatively high average of accuracy, meaning they thought they did worse than they actually did. Finally, for the novice's group, the visualization that had the highest average of efficacy was the fifth visualization, which again may be due to the fact that the idiom was a bar chart and the participants may have felt more at ease with this type of visualization.

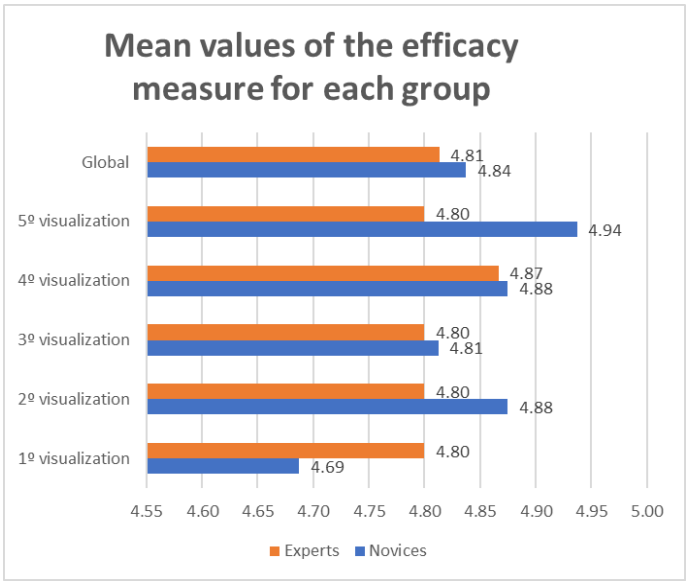


Figure 12 - Mean values of the efficacy measure for each group

As for the experts group, all of the visualizations obtained the same level of efficacy, with the exception of fourth visualization, which in turn was not one of the visualizations that obtained one of the highest averages of accuracy, meaning that the experts group may have perhaps thought they did better on this visualization than they actually did and they may have thought they had the same level of performance on the other visualizations.

Another measure that was also analysed was the engagement, which can be understood as the level in how the participants would classify their desire to interact with each visualization. The results of this measure can be seen in **Figure 13**, and the first thing that is noticeable is that there is a more significant difference in the global results of the novices group and the experts group, where the novices group has a relatively higher average of

engagement with a value of 4.83 and the experts group has a value of 4.44. It is also worth noticing that all the results were above 4, which means that again they classify as the fourth level of High engagement in both groups. The group of novices also has higher values of engagement than the expert's group in all of the visualizations.

The visualizations with the lowest average of engagement for the experts group are the fifth and first visualization, which again correspond to the bar chart and the stacked bar chart idioms that may be a very common type of visualizations for the experts group and therefore does not create as much engagement as other visualizations. On the contrary, the visualization that had the highest average of engagement for the experts group was the second visualization and although there were other visualizations with more interactive features than the second one, these results could be due to the fact that it was a choropleth map, which may not be a very common type of visualization and therefore increased the engagement.

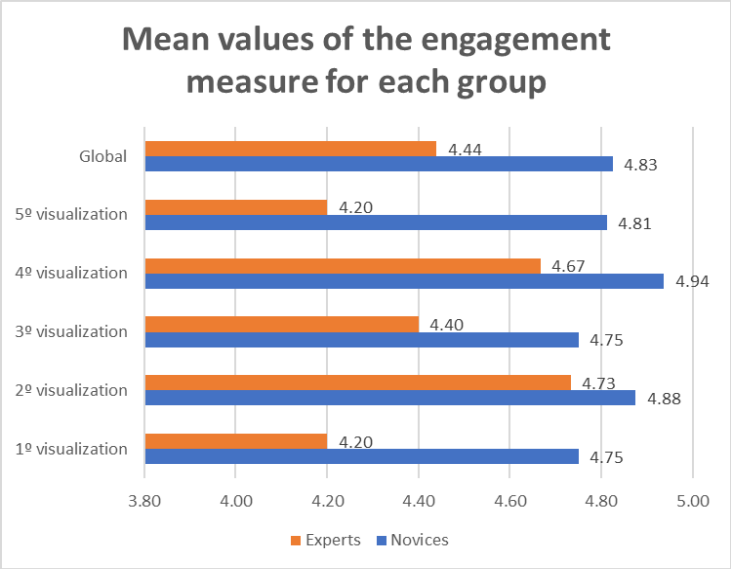


Figure 13 - Mean values of the engagement measure for each group

As for the novices group, the visualizations with the lowest average of engagement were the third visualization, where the idiom was an heatmap, and the first visualization, where the idiom was an stacked bar chart, maybe due to the fact that they only had two interactive features, the select box and the tooltips for the case of the third visualization and the slider and the tooltips for the case of the first visualization. Regarding the heatmap it could also have been because the heatmap may be more intimidating to analyse, especially for the novice's group and that could have decreased the engagement that the participants had with the visualization. The visualization that had the highest average of engagement was the

fourth visualization, probably due to the fact that it had 3 interactive features – slider with a play button that provided an animation, radio box and tooltips - which is the maximum that a visualization had in the prototype, and that likely increased the engagement with the visualization.

Finally, the last main measure to be analysed was the complexity, which could be understood in how hard the participants thought it was to analyse each visualization. The results can be observed in **Figure 14**, and the main observation is that the experts group had global average of complexity slightly higher than the novices group, with a value of 1.96 and the novices group with a value of 1.87, meaning that for both groups the average complexity classified as Very low. Another interesting fact is that in each visualization, both groups had always the same interval classification for the visualizations, where they classified the first, the third and the fourth visualization with Low complexity, where the idioms were a stacked bar chart, an heatmap and a scatterplot chart and the second and the fifth with Very low complexity, where the idioms were a choropleth map and a bar chart. It is also worth mentioning, that the expert's group usually had a slightly higher average of complexity than the novice's group, with the exception of the third visualization where they had the same average values and the fifth visualization where the novices group had higher values than the expert group.

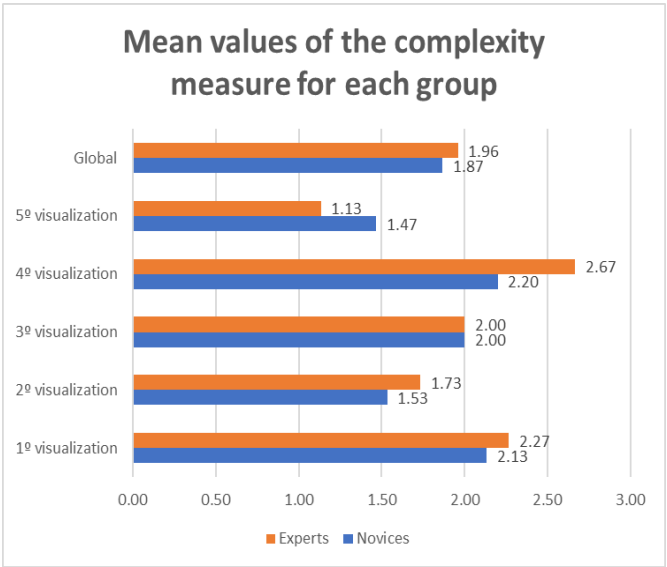


Figure 14 - Mean values of the complexity measure for each group

The visualization that had the highest average of complexity either for the group of experts or the group of novices, was the fourth visualization, maybe due to the intricacy of the

scatterplot chart idiom and due to increased number of interactive features. On the opposite side, the visualization that had the lowest average of complexity, also for both types of groups, was the fifth visualization, maybe due to the fact that, as mentioned previously, the bar chart is known as a more common type of visualization and it may be easier to analyse, even for the novices group.

As mentioned in the beginning of this section, the measures of complexity and usefulness were also analysed for all the components by calculating the global average of each measure for all the components of each visualization and then comparing the visualizations and their global value. The first measure analysed was the usefulness and it is possible to see the results of this analysis in **Figure 15**.

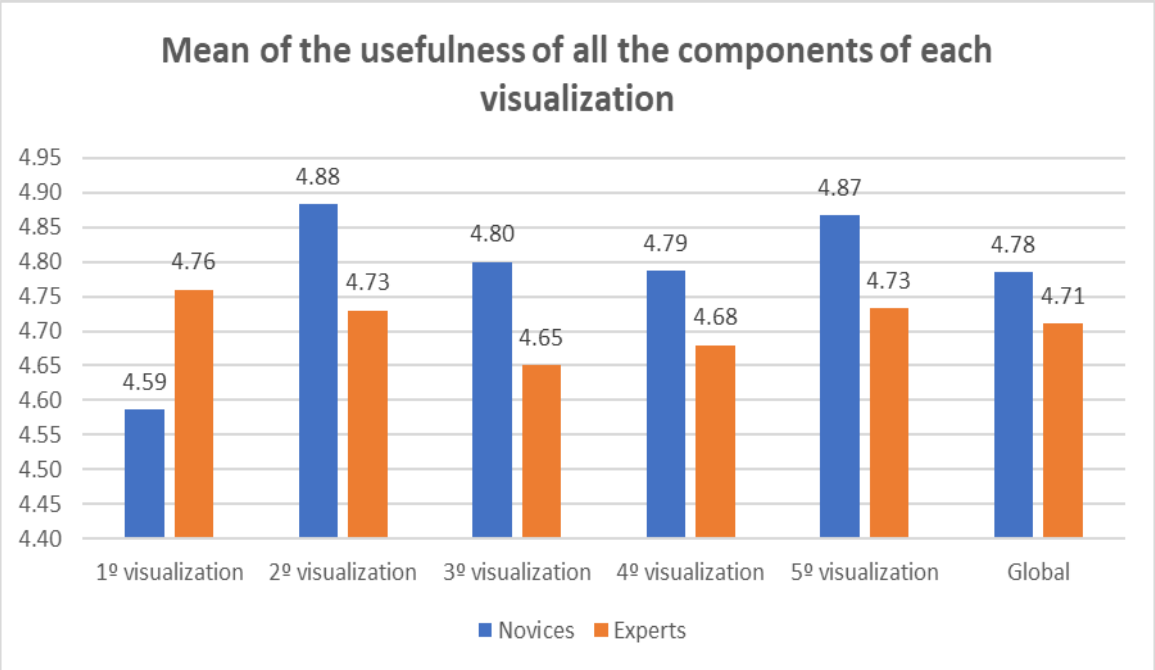


Figure 15 - Mean of the usefulness of all the components of each visualization

One of the first observations is the fact that the novice's group had an average of usefulness of the components higher than the expert's group, with a value of 4.78 and the expert's group with a value of 4.71. This is coherent with the results of the usefulness of the prototype and the visualizations in general. Same with previous analysis, the novice's group has higher averages of usefulness for the components than the expert's group, with the exception of the first visualization. It is also worth noticing that, similar with previous results, all the values are above 4.0 which means that, on average the components of all the visualizations were classified as High in terms of usefulness.

The first visualization is also the visualization with the lowest average of usefulness for the components for the novice's group, being again, coherent with the results of the analysis of the usefulness measure in general. The visualization with the highest average of usefulness for the components for the novice's group is the second visualization, which also obtained a higher average of usefulness in the previous analysis.

As for the expert's group the visualization that had the lowest average of usefulness of the components was the third visualization, which is interesting considering this was the visualization where the experts group obtained the highest average of usefulness in general and the highest average of accuracy. A possible explanation for this may be the fact this visualization only contains two interactive features, the select box and the tooltips, and that the participants may have thought that the components weren't as useful as in other visualizations. Regarding the visualization with the highest average of usefulness of the components for the expert's group, that was the first visualization where the idiom chosen was the stacked bar chart. In **Table 9** it is possible to observe which components had the highest and lowest average of usefulness in each visualization.

Table 9 - Components with the highest and lowest average of usefulness for each visualization

Usefulness	**Highest**		**Lowest**	
Visualizations/ Types of groups	**Novices**	**Experts**	**Novices**	**Experts**
1º visualization	Stacked bar chart 4.8	Slider and tooltips 4.87	Text 4.33	Text 4.53
2º visualization	Choropleth map 5.0	Choropleth map and slider 4.93	Text 4.73	Text 4.27
3º visualization	Heatmap and tooltips 4.87	Select box and tooltips 4.87	Text 4.67	Text 4.13
4º visualization	Slider 4.93	Radio box 4.87	Text 4.60	Text 4.20
5º visualization	Bar chart, slider and radio box 5.0	Radio box 4.93	Text 4.40	Text 4.27

With the exception of the fourth visualization for the novice's group, the component that had the highest average of usefulness is usually the visualization itself, which is logic considering it is the main part of the analysis. In general, the novice's group gives higher values of usefulness to the components than the experts group, which is coherent with the previous results. For the novices group the components with the highest average of usefulness were the choropleth map in the second visualization and the bar chart, slider and radio box in the fifth visualization with a maximum average of 5.0, similar as the expert's group where the components that had the highest average of usefulness were the choropleth map and the slider in the second visualization and the radio box in the fifth visualization with a value of 4.93, meaning that the most useful components for both groups were in the second and fifth visualization. The least useful component for both groups in all the visualization was the text, which has logic considering the text present on the visualizations was there to help clarify the ideas in the visualization but was not the main part of the analysis and was also the only feature that was not interactive. It is also interesting to see that the tooltips, the slider and the radio box are also interactive features that had the highest averages of usefulness in several visualizations, either for the group of experts or for the group of novices. The only exception was the select box which was only considered to be one of the most useful components in the expert's group, meaning that it might not be as useful for the novice's group.

Besides the usefulness, it was also analysed, as previously mentioned, the average complexity of the components of each visualization and of the prototype in general. The results of this analysis can be seen in **Figure 16**.

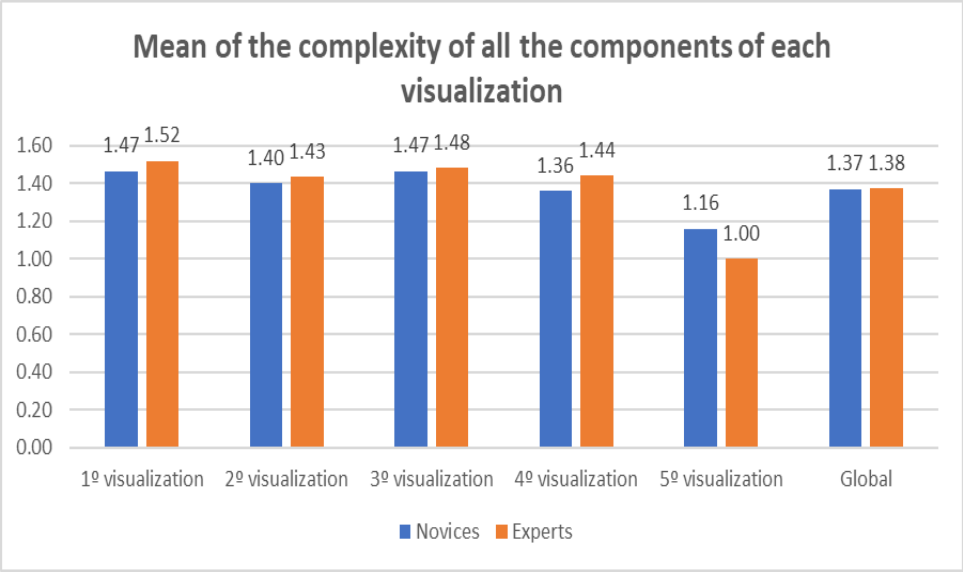


Figure 16 - Mean of the complexity of all the components of each visualization

Same as the complexity measured analysed in general, in the complexity of the components, the expert's group has a slightly higher average than the novice's group, with a value of 1.38 and the novice's group with a value of 1.37. Although, the expert's group has slightly higher averages of complexity of components in most of the visualizations, with the exception of the fifth visualization, the gap between the two groups is minimal, and all of the average values are under 2.0, meaning they are all classified as Very low complexity. It is interesting to notice that, the only visualization where the expert's group had a lower average complexity of the components than the novice's group was the fifth visualization, perhaps due to the fact that the expert's group is even more familiarized with bar charts than the novice's group. The fifth visualization was also the visualization that had the lowest value of complexity regarding the components, for both groups of individuals, being coherent with the results of the analysis of the complexity measure analysed in general.

The visualization with the highest average of complexity regarding the components for the expert's group was the first visualization, possible due to the text, since it is the visualization with the highest density of text and that can be a complex component. Another possible reason may be because the stacked bar chart contains an abundance of information with several different variables and since it was the first visualization that the participants were exposed too, that may have caused some difficulties that therefore increased the average of complexity of the components for the expert's group.

As for the novice's group, the visualizations that had the highest average of complexity were the first visualization, possible due to the same reason as the experts group, and the third visualization, where the idiom was an heatmap, which again may not be a very common type of visualization and the participants were probably not as familiarized with this type of visualizations as they were with other types, increasing the average of the complexity of the components, since the type of idiom itself is a component of the visualization as a whole. In **Table 10** it is also possible to see which components had the highest and lowest average of complexity in each visualization.

As for the component with the highest average of complexity in the novice's group, those were the heatmap in the third visualization - which is coherent with the previous results considering that one of the visualizations that had the highest average of complexity for the components in the novice's group was the third visualization - and the scatterplot chart in the fourth visualization. A possible reason to this may be the fact, that as previously mentioned,

neither the heatmap nor the scatterplot chart are usually considered to be very common types of visualization and that may have caused some difficulties for the participants when interacting with this type of idioms. The component that had the highest average of complexity in the expert's group was the heatmap in the third visualization, possible due to the same reason as the novice's group. If the actual visualization is not included in this analysis, the component that has the highest complexity for the novice's group is the text and the radio box and for the expert's group is the select box.

Table 10 - Components with the highest and lowest average of complexity for each visualization

Complexity	**Highest**		**Lowest**	
Visualizations/ Types of groups	**Novices**	**Experts**	**Novices**	**Experts**
1º visualization	Stacked bar chart and the text 1.67	Stacked bar chart 1.87	Slider 1.13	Slider 1.20
2º visualization	Text 1.60	Select box 1.73	Slider and select box 1.27	Slider 1.13
3º visualization	Heatmap 1.73	Heatmap 1.87	Text 1.33	Tooltips 1.27
4º visualization	Scatterplot chart 1.73	Scatterplot chart 2.07	Tooltips 1.13	Slider and text 1.20
5º visualization	Radio box 1.27	Same for all components 1.0	Slider 1.07	Same for all components 1.0

Regarding the component with the lowest average of complexity, for both groups this component was present in the fifth visualization, and for the novice's group it was the slider with a value of 1.07 and for the expert's group it was all the components of that visualization, with a value 1.00 for all of them. This is coherent with the previous results, since as previously mentioned, the bar chart is usually known as a more common type of visualization and the participants may have felt more comfortable with this type of idiom, decreasing their complexity. Besides that, the slider was also referred several times as the component with the lowest average of complexity in several visualizations for both groups, meaning that both

groups felt comfortable interacting with this type of feature, whereas the radio box was one of the components with the highest average of complexity in the fifth visualization for the novice's group, which can indicate that this feature may be harder to interact than other types of interactive features.

The tooltips were also mentioned as one of the least complex features, possible meaning that they are also an interactive feature that both groups did not have difficulties to interact with it. Lastly, the text was seen as both as one of the most complex components and one of the least complex, meaning that this component may vary in complexity depending on the way it is applied.

Regarding the qualitative measures, the only results that are left to be analysed are the average rankings of each visualization for each group, which were asked at the end of the questionnaire "Qualitative questions". The results of this analysis for the novice's group can be seen in **Figure 17**, where the results for the expert's group can be seen in **Figure 18**.

It is interesting to notice that for both groups the order of rankings is the same, where the visualization that ranked in first place is the third visualization where the idiom is an heatmap, which is interesting considering is not a very common type of idiom and although it was the visualization that had the highest average of accuracy and lowest completion times, it also was one of the visualizations that had an higher average of complexity regarding the components for both groups. On the opposite side, the type of idiom that ranked in last was the choropleth map for both groups.

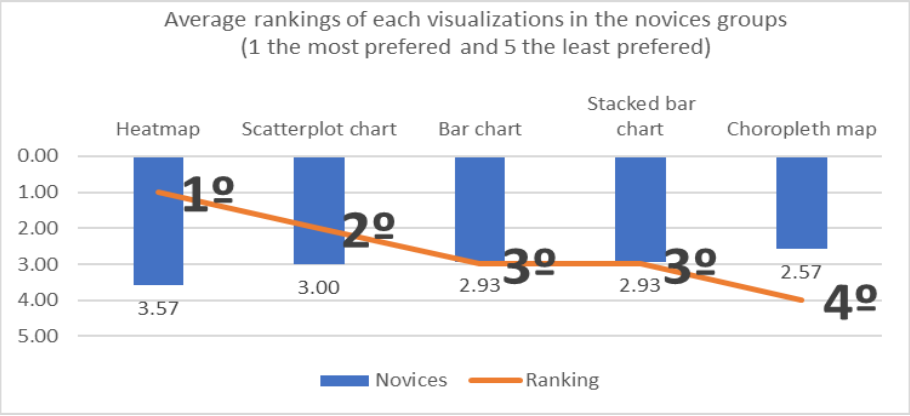


Figure 17 - Average rankings of each visualization in the novice's groups

Another interesting observation is that for the novice's group the bar chart and the stacked bar chart ranked in the same position and obtained the same average ranking of 2.93, which has logic considering they are similar types of idioms.

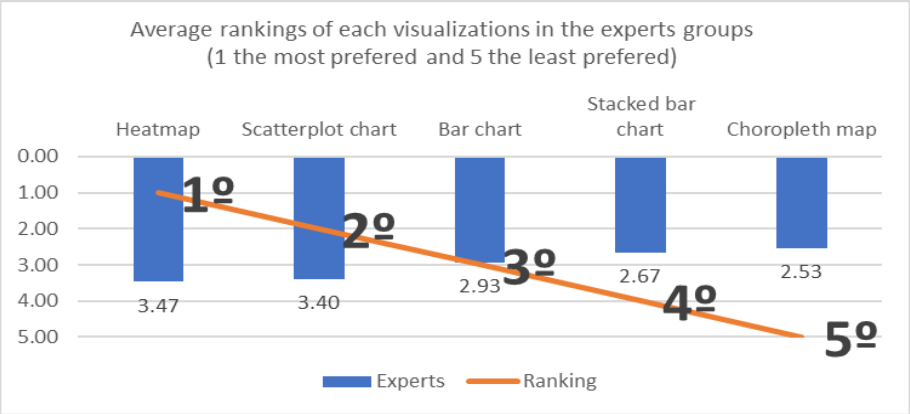


Figure 18 - Average rankings of each visualization in the expert's groups

Regarding the rankings, a final observation is that the novice's group had an higher average ranking for the third visualization – where the idiom was an heatmap - than the expert's group, with a value of 3.57 compared to 3.47, possibly meaning that the novice's group enjoyed the heatmap even more than the expert's group, and that the expert's group had a lower average ranking for the second visualization – where the idiom was an choropleth map – than the novice's group, with a value of 2.53 compared to 2.57, meaning that the expert's may have disliked the choropleth map more in comparison with other visualizations, than the novice's group.

Finally, it was also analysed as whole, the final averages of all four measures for both groups, usefulness, efficacy, engagement and complexity in order to more easily compare both groups. These results were summarized in **Figure 19**.

On conclusion, it can be observed that on average, the expert's group ranked the components of each visualizations, the visualizations and prototype as whole, with a higher average of complexity than the novice's group. On the contrary the novice's group had a higher average of usefulness, – either for the prototype and the visualizations or for the components of each visualization – efficacy and engagement than the expert's group.

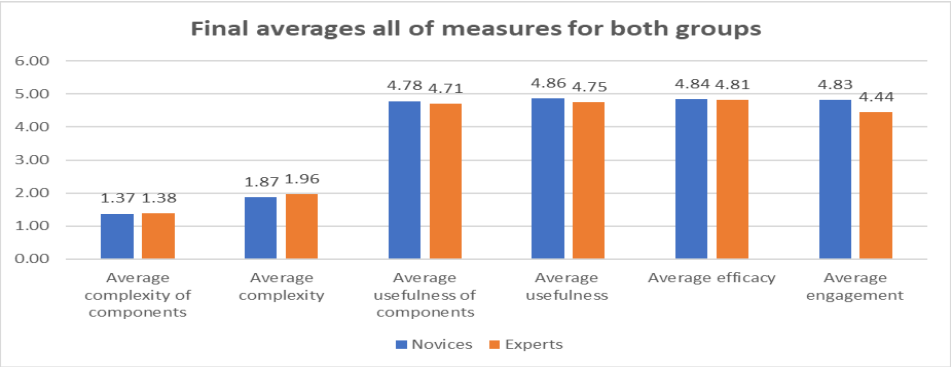


Figure 19 - Final averages of all qualitative measures for both groups

Regarding the qualitative measures, an ANOVA single factor test was also conducted to test if the differences found between both groups were statistically significant. For the usefulness measure the results did not have differences that were statistically significant ($F(1,8) = 2.119055$, $p = 0.183568$), as it is possible to see in **Table 11**. The same applied for the usefulness measure regarding the components ($F(1,8) = 1.717874$, $p = 0.22634$), which can be seen in **Table 12**.

Table 11 - ANOVA single factor test for the usefulness measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	24.32	4.864	0.02423
Experts	5	23.74	4.748	0.00752

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.03364	1	0.03364	2.119055	0.183568	5.317655
Within groups	0.127	8	0.015875			
Total	0.16064	9				

Table 12 - ANOVA single factor test for the usefulness of the components measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	23.92333	4.784667	0.013976
Experts	5	23.5531	4.710619	0.001983

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.013708	1	0.013708	1.717874	0.22634	5.317655
Within groups	0.063835	8	0.007979			
Total	0.077543	9				

For the complexity measure the same results were also obtained, where no statistically significant differences were found, neither for the complexity in general ($F(1,8) = 0.0975$, $p = 0.762835$), as can be seen in **Table 13**, or for the complexity of the components ($F(1,8) = 0.001779$, $p = 0.967388$), that can be seen in **Table 14**.

Table 13 - ANOVA single factor test for the complexity measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	9.33	1.866	0.11723
Experts	5	9.8	1.96	0.3359

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.02209	1	0.02209	0.0975	0.762835	5.317655
Within groups	1.81252	8	0.226565			
Total	1.83461	9				

Table 14 - ANOVA single factor test for the complexity of the components measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	6.853333	1.370667	0.015947
Experts	5	6.876667	1.375333	0.045253

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	5.44E-05	1	5.44E-05	0.001779	0.967388	5.317655
Within groups	0.2448	8	0.0306			
Total	0.244854	9				

Finally, regarding the efficacy measure, again no statistically significant differences ($F(1,8) = 0.333662$, $p = 0.579401$) were found between both groups, as can be seen in **Table 15**. However, as for the engagement measure, this was the only measure where statistically significant differences ($F(1,8) = 10.58511$, $p = 0.01164$) were found, considering the p value was under 0.05, as can be seen in **Table 16**. This means this was the only qualitative measure were the differences between both groups were statistically significant.

Table 15 - ANOVA single factor test for the efficacy measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	24.2	4.84	0.00915
Experts	5	24.07	4.814	0.00098

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.00169	1	0.00169	0.333662	0.579401	5.317655
Within groups	0.04052	8	0.005065			
Total	0.04221	9				

Table 16 - ANOVA single factor test for the engagement measure

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Novices	5	24.13	4.826	0.00693
Experts	5	22.2	4.44	0.06345

ANOVA

Source of variance	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between groups	0.37249	1	0.37249	10.58511	0.01164	5.317655
Within groups	0.28152	8	0.03519			
Total	0.65401	9				

4.2.3. Experiment observations

In this section it will be analysed the main experiment observations collected in both stages of the experiment of this project. The experiments observations were divided into two main categories: suggestions, where some of them were implemented in the next stage of the prototype improvement, and observations, which were remarks or notes that the participants made throughout the experiment and that were also taken into consideration for the prototype improvement.

Due to the extensiveness of the data collected in the experiments of this study, the complete list of suggestions can be seen in **Table 20** in **Annex C,** where it is possible to see

all the suggestions made in the experiments and the number of participants that also made that suggestion in each group of individuals, plus if that suggestion was made by individuals of both groups. The same information is also in the complete list of observations that is present in **Table 21** in **Annex D**.

The main conclusions that were taken from this analysis were that, regarding the suggestions, the most common suggestions for the group of novices were that the play button could be in a lower position, with five participants supporting this suggestion and that the map of the second visualization could have values and a legend for each country, with four participants supporting this suggestion. Both of these suggestions were also some of the most popular suggestions regarding the expert's group, as well as suggestions such as: making the format of tooltips more constant in all the visualizations and increasing the size of the letters, which are also suggestions that the participants from the novice's group also made.

Other popular suggestions from the novice's group were also: the fact that the slider in the first visualization should not allow for only one year to be selected; that it would be simpler in the cases of specific years to just have a box to enter the year wanted, instead of selecting a specific year in the slider (without the play) and only allow the use of the slider with play; and finally the suggestion to explain in the text of the fourth visualization what is the variable weight, to facilitate the analysis. All of these suggestions had two participants from the novice's group supporting each decision.

Regarding the expert's group, other suggestions that were also considered to be popular were: applying the same logic to all visualizations considering the third visualization is different than others because it doesn't have a slider and the time analysis is more difficult; presenting the text of the first visualization in a different format or decreasing the size of it, and finally, highlighting Portugal in the labels of the fifth visualization so it can be visible even if it has null values. All of these suggestions had two participants that supported them, with the exception of the suggestion to put the format of the tooltips constant in all the visualizations, which had three participants supporting that suggestion, being that the maximum of participants from the expert's group that agreed on a suggestion.

Other suggestions that were also supported by individuals of either the group of experts or the group of novices, were suggestions such as: the fact that the play button could be more prominent; the fact that the heatmap could have less colour variance to improve

visibility, the decrease of the speed of the animation of visualizations; the fact that in the second visualization the map could have a zoom and could be aggregated for economical geographical areas; the use of ISO Codes 2 in the fifth visualization to facilitate the analysis; and finally the fact that in the third type of visualization the items in the select box should be organized by all goods, all services and all current and capital account instead of organizing by all exports, all imports or all balances.

Regarding the observations, the most common observation made, either in the group of experts or in the group of novices, was that the play button used in some visualizations was very engaging and intuitive, where eight participants from the novices group supported this observation and eleven from the expert's group. Besides this observation, some other remarks that were also common to individuals from the two different types of groups were: for some participants the select box was more useful than the radio box, with four participants from the novice's group supporting this observation and two from the expert's group; the radio box was more intuitive than the select box because it was possible to see all the choices, with seven participants from the novice's group supporting this observation and five from the expert's group and finally, the radio box was more intuitive than the select box but only if it had few categories, since if it had a lot of categories the select box was preferred, with four participants from the novice's group supporting this decision and six from the expert's group. This means that the participants from both groups were very indecisive regarding which interactive feature was the better, the select box or the radio box, so the decision of which component might be better for each group will be decided by the results of the qualitative measures regarding the components.

Some other interesting observations, that were also common to both types of groups, was the fact that some participants preferred to use the slider with the play button, with four participants from the novice's group agreeing with this observation and eight participants from the experts, whereas some participants, preferred the slider without the play button if it was for a specific year but to show changes and in general preferred the slider with the play, where three participants from the novice's group also made this observation and two participants from the expert's group. A final observation that is worth taking into consideration is the fact that there were participants from both groups that mentioned that the heatmap was the easiest map to analyse and their favourite visualization, where three of them were participants from the novice's group and two from the expert's group. This can be

coherent with the results from the previous analysis, considering that in the rankings the heatmap had the highest average ranking in both groups.

4.3. PROTYPE IMPROVEMENT

As mentioned previously, the final step of the methodology flowchart applied in this project was the conclusion phase, where using all the feedback collect from the experiments, a list of improvements was applied to the project in order to improve and finish the prototype of this study.

Using the experiment observations, it was possible to analyse 52 suggestions made by the participants of both groups, in the experiments. Out of these 52 suggestions, 26 were implemented. The remaining ones were not implemented due to one (or more) of these possible reasons:

- The impossibility of implementing said suggestion, due to the constrains of the software and packages used, as well as lack of resources, considering the fact that some suggestions would require an excel level of skill of the R programming language and all the including packages (*Reason 1*);

- Implementing said suggestions would change the visualization so much that it would require a new analysis with a completely different experiment or the fact that said suggestion would not make sense to the analysis in question (*Reason 2*);

- It would only be relevant for the economy analysis, which is not the focus of this study, and it would not be valid for the visualization component of the prototype (*Reason 3*).

In **Table 20** in **Annex C** it was also added an extra column, stating whether or not the suggestion was implemented, and which was the reason not to implement.

Regarding the first visualization, the suggestion that was implemented only to this visualization was: changing the colours of the stacked bar chart to more intuitive colours, changing the primary income to green and the secondary income to blue, as suggested. The first visualization after modifications can be seen in **Figure 20**.

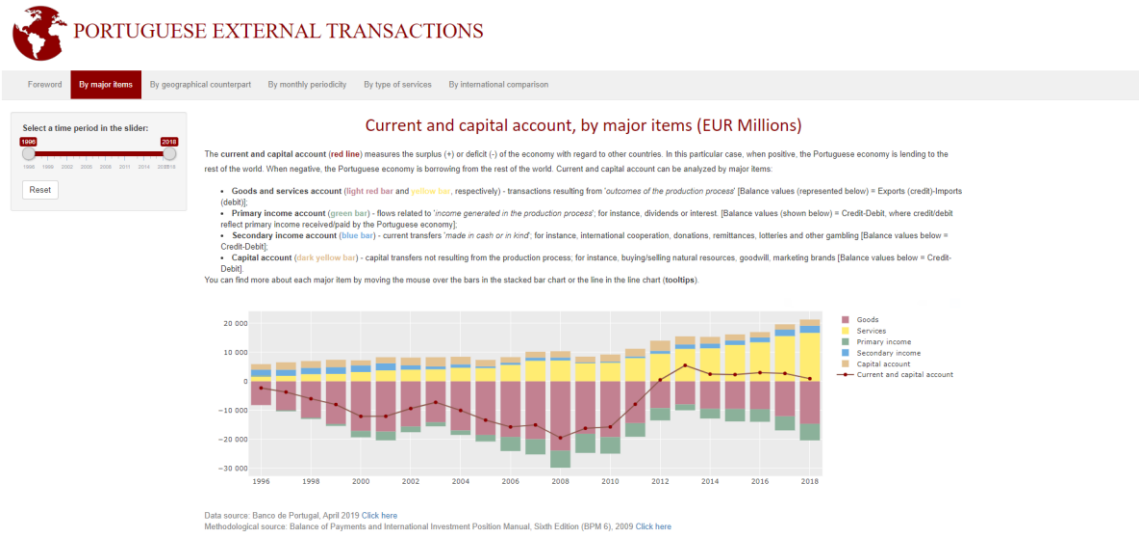


Figure 20 - Prototype "By major items" tab after modifications

As for the second visualization, the main suggestions implemented were: changing the colour pallet of the choropleth map to an orange colour palette, as suggested, since the colours could be more distinguishable; putting in the text the section related to the quartiles in bold; mentioning in the text the definition of quartiles, and explaining in the text that the visualization is about Portugal with the rest of the world. The second visualization after modifications can be seen in **Figure 21**.

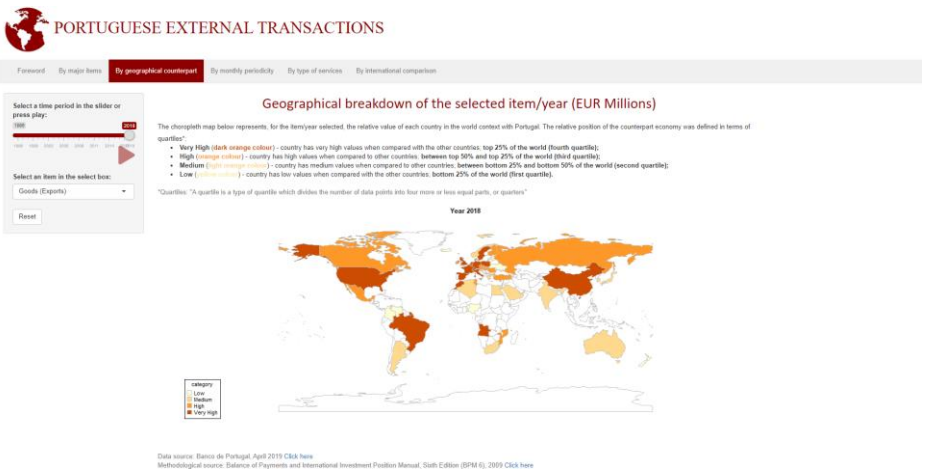


Figure 21 - Prototype "By geographical counterpart" tab after modifications

The suggestions that were implemented only in the third visualization were: organizing the items in the select box by all goods, all services and all current and capital account, instead of organizing by all exports, all imports or all balances; putting the select box in a lower position parallel with the heatmap; increasing the size of the heatmap to the maximum of its width to decrease the probability of error, and decreasing the amount of categories in the colour palette of the heatmap to reduce the colour variance, in order to facilitate analysis. The third visualization after modifications can be seen in **Figure 22**.

In the fourth visualization, the main suggestions that were implemented were: highlighting the travel and transport services with the use of a border instead of colour, in order to maintain the colour in the rest of the services; mentioning in the text that there is an option to see all the types of services displayed or all the types of services excluding the transport and travel; putting the section that mentions the bissectrice in the text in bold; changing some of the colours that categorized the services, considering that there were two shades of blue that were too similar, and explaining in the text the meaning of the variable weight, mentioned in the tooltips. The fourth visualization after modifications can be seen in **Figure 23**.

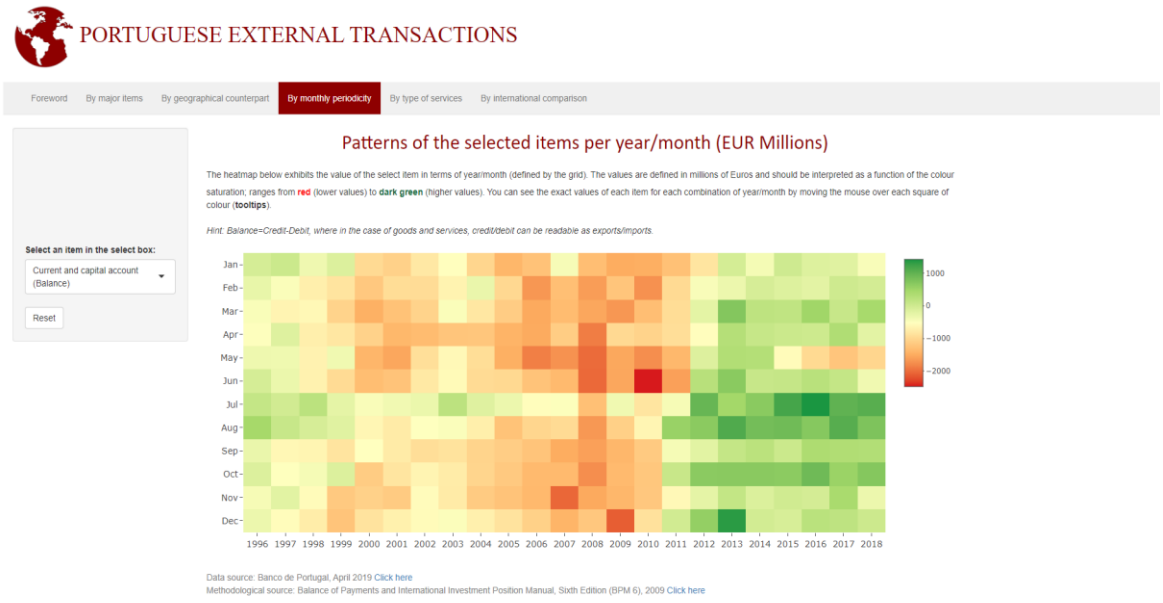


Figure 22 - Prototype "By monthly periodicity" tab after modifications

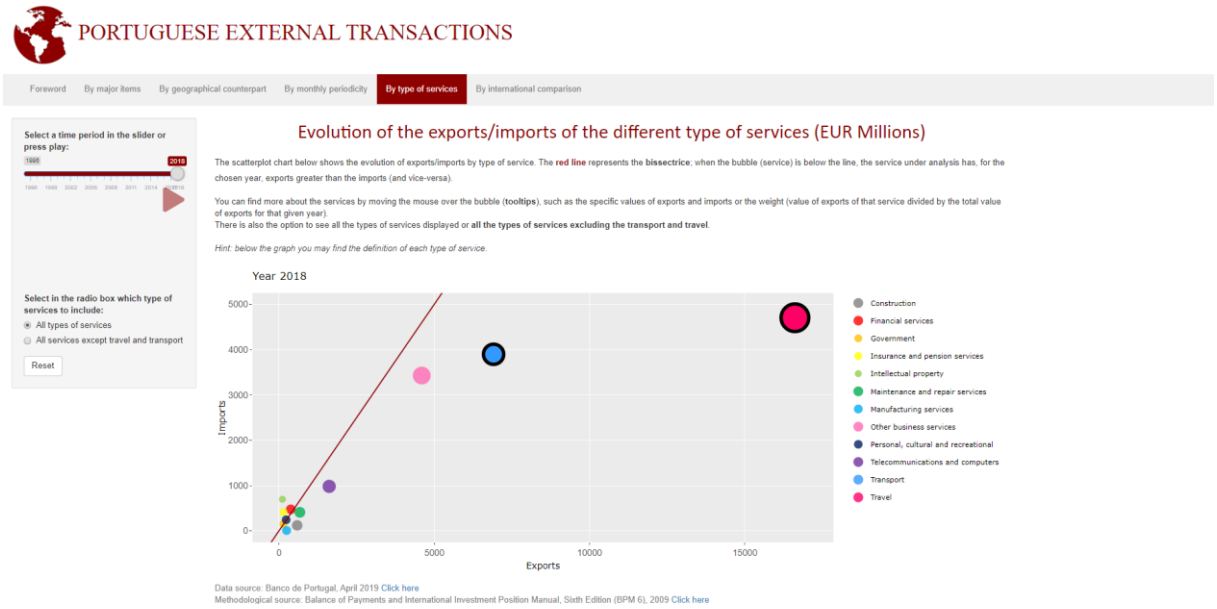


Figure 23 - Prototype "By type of services" tab after modifications

Finally, the only suggestions that were exclusive to the fifth visualization were: mentioning the full name of the country in the tooltips, instead of just the ISO Codes; changing the labels from ISO Code 3 to ISO Code 2, as suggested; increasing the size of the fifth visualization, making it easier to analyse; make the y axis dynamic so it changes throughout the years; changing the colours of the bars of the other countries besides Portugal, from yellow to grey, since yellow and red (colour of the Portugal bar) were colours in the same range of temperature, and lastly, changing some parts of the text to facilitate the analysis. The fifth visualization after modifications can be seen in **Figure 24**.

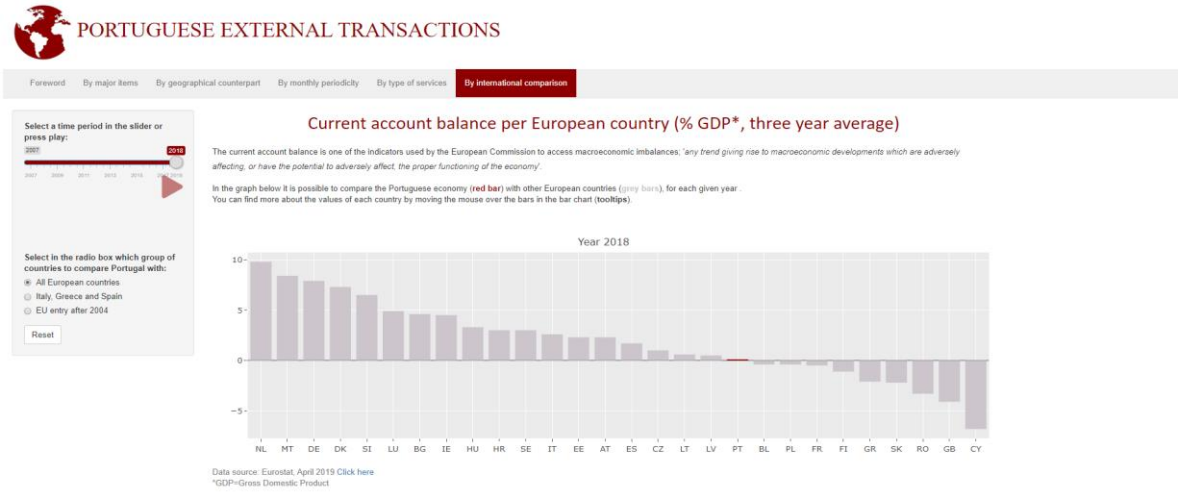


Figure 24 - Prototype "By international comparison" tab after modifications

There were a few suggestions that were implemented for several different visualizations, such as: increasing the size of the play button, so it would be more prominent, or decreasing the speed of the animations, in the second, fourth and fifth visualization; placing ticks in the all of the sliders, including the first, second, fourth and fifth visualization; changing the format of the tooltips to a more constant format in all of the visualizations that possessed them, which were the first, third, fourth and fifth visualizations; putting the radio boxes of the fourth and fifth visualization in a lower position in parallel with the visualizations in order to highlight them, and finally, create a reset button in all of the visualizations of the prototype.

Regarding the remaining suggestions, some of the most popular ones that were not implemented were: not putting the player button in a lower position, due to constrains of the slider widget of the Shiny package; not adding tooltips with the values of each country in the choropleth map of the second visualization, considering it would have required a change of package - since the package used to develop the choropleth map *rworldmap* does not have the possibility to add this feature – and that would have changed the visualization entirely; not increasing the size of the letters, since in some visualizations it could affect the rest of the visualization and their idioms, and it would not be possible to see them in just one screen without having to scroll down; not applying the same time analysis logic to all the visualizations, considering it was necessary to test different types of visualizations, and some idioms required different approaches than others; not changing the text of the first visualization, because although it was a very extensive text compared to others, considering there are two types of groups, experts and novice's, it was necessary to explain the visualization in further detail to facilitate the analysis to both groups, and adding a zoom or an aggregation for economical geographical areas in the choropleth map of the second visualization, since that would also have required a different package than the one used, and as previously mentioned, that would have changed the visualization entirely.

There were also three other popular suggestions (had more than one participant supporting that suggestion), that were not implemented due to first reason mentioned above, which means due to the constraints of the software used and lack of experience with the programming language. Those three suggestions were: not making the slider in the first visualization not allow the selection of only one year; highlighting Portugal in the labels of the fifth visualization so it could be visible even if it had null values, and implementing the

suggestion of in the case of the specific years for the sliders, only having a box to enter the year wanted, instead of selecting a specific year in the slider (without the play), therefore only allowing the use of the slider with the play.

Finally, it is also worth mentioning that in the Foreword page no suggestions were implemented, since the participants did not mention any suggestions that involved that first tab, so it remained the same.

5. DISCUSSION

In this section of this chapter, it will be performed a critical analysis of the results obtained from the experiments of this project, as well as comparison of these results with results from other studies, mentioned in the literature review.

As mentioned previously, the main gap this study was going to try to address was to understand the differences between how different types of individuals discoverer interactivity, and which are the most appropriate visualization techniques for each type of individual. In addition to this, another goal was to discover if an open tool, such as the Shiny package of the R programming language was an adequate tool to develop interactive visualizations for different types of individuals.

The main reason this study was performed was the fact that, as mentioned previously, data visualization is important in a broad set of fields, and therefore if companies or organizations that use it in their business, start to understand the differences between how different types of individuals discover interactivity and which are the best visualization techniques for each group, that will likely give them a competitive advantage since they will more efficiently adapt their visualizations to each type of individuals.

This study contributed with a methodology to evaluate the differences between experts and novices. The prototype was tested with two different groups of individuals to discover which were the best visualization and interaction techniques to use for the different groups. In addition, this study also considered the creation of a prototype using the Shiny package in the R programming language, which was a different tool than the other previous studies have used, with the exception of the work conducted by López *et al.* (2018) that had a more restricted goal since it only tried to test the efficacy of the Shiny tool with students, or the works conducted by Seal & Wild, (2018) and Su, Sun, Shimizu, & Kadota (2019), that did not have the objective of testing the differences in how two different groups discover interactivity.

The overall approach for studying the gap of this study was to create a prototype using the Shiny package in the R programming language, with different interactive visualizations and in each of them using different interactive and visualizations techniques, using data related with the economy field, from *Banco de Portugal*, contrarily to previous studies (Oghbaie, Pennock, & Rouse, 2016; Rouse, Pennock, Oghbaie, & Liu, 2017). To create the

prototype, it was also used the methodology flowchart that, as mentioned previously, took inspiration in the design and creation process suggested by Oates (2005). After the creation of the prototype, it was then performed the validation, using an experiment with 30 participants of two different types of groups, experts and novices. To perform the validation of the prototype, quantitative measures were used, such as accuracy and completion time, and qualitative measures, such as the usefulness, the complexity, the engagement and the effectiveness.

The experiments performed also developed results regarding all the metrics analysed, which can be seen in **Table 17**. To test for statistically significant differences an ANOVA single factor test was used. Regarding the quantitative measures no statistically significant differences between both groups were found. As for the qualitative measures although it appeared that there were differences between groups, the engagement was the only metric where these differences were statistically significant. This means that this is the only metric where results can be improved in order to close the gap between the group of experts and novices, and perhaps if only the best visualizations for each group were used, this gap could have been closed.

Table 17 - Results of the metrics analysed for the different groups

Metrics	Novices	Experts	ANOVA results
Accuracy	**91%**	**93%**	**(F(1,8) = 0.122605, p = 0.735266)**
Time	**21 minutes and 57 seconds**	**24 minutes and 1 second**	**(F(1,8) = 0.358601, p = 0.565851)**
Usefulness	**4.86**	**4.75**	**(F(1,8) = 2.119055, p = 0.183568)**
Usefulness of the components	**4.78**	**4.71**	**(F(1,8) = 1.717874, p = 0.22634)**
Efficacy	**4.84**	**4.81**	**(F(1,8) = 0.333662, p = 0.579401)**
Engagement	**4.83**	**4.44**	**(F(1,8) = 10.58511, p = 0.01164)**

Complexity	**1.87**	**1.96**	**(F(1,8) = 0.0975, p = 0.762835)**
Complexity of the components	**1.37**	**1.38**	**(F(1,8) = 0.001779, p = 0.967388)**

In **Table 18** it is possible to observe which were the best and worst visualizations for each group. Regarding the best visualization for each group it is possible to conclude that for the novice's group the best visualizations were the third, where the idiom was an heatmap, which was the visualization that ranked in first place and had the highest values of accuracy and also had one of the lowest completion times, and the fifth visualization, where the idiom was a bar chart, since it was the visualization with the lowest completion time, the highest average of usefulness, the highest average of efficacy and the lowest average of complexity either in general or regarding the components.

As for the expert's group, the best visualization was also the third visualization, for the same reason as the novice's group and since it had the highest average of usefulness and the lowest completion time. Besides the third visualization, for the expert's group it was not as simple as the novice's group considering the results were more spread throughout the visualizations, where the fourth visualization had the highest average of efficacy, the second visualization had the highest value of engagement, the fifth visualization the lowest value of complexity in general and of all the components, and the first visualization had the highest value of usefulness for the components. Although the fifth visualization also had relatively lower averages of usefulness and engagement when compared to other visualizations, considering that the fifth visualization had the best averages in two qualitative measures and it had some of the highest averages of accuracy and some of the lowest averages of completion times, it is also possible to consider that for the expert's group the best visualizations are the third visualization and the fifth, meaning that there are no differences between the two groups in which were the best visualizations.

As for the worst visualizations for each group, the conclusions were that for the novice's group it was the second visualization, where the idiom was a choropleth map, considering that it was the visualization that ranked in last and had the lowest accuracy average, and the first visualization, where the idiom was an stacked bar chart, considering it had the highest completion time, the lowest averages of usefulness in general and of the

components, efficacy and engagement and also had one of the highest averages of complexity in general and of the components.

Table 18 - Best and worst visualizations for each measure for the different groups

Measure Analysed	Visualization with the best values in the group of experts	Visualization with the best values in the group of novices	Visualization with the worst values in the group of experts	Visualization with the worst values in the group of novices
Accuracy (Highest better)	Heatmap (3rd visualization) and Bar Chart (5th visualization)	Heatmap (3rd visualization)	Stacked bar chart (1st visualization)	Choropleth map (2nd visualization)
Completion time (Lowest better)	Heatmap (3rd visualization)	Bar Chart (5th visualization)	Choropleth map (2nd visualization)	Stacked bar chart (1st visualization)
Complexity (Lowest better)	Bar Chart (5th visualization)	Bar Chart (5th visualization)	Scatterplot chart (4th visualization)	Stacked bar chart (1st visualization)
Complexity of components (Lowest better)	Bar Chart (5th visualization)	Bar Chart (5th visualization)	Stacked bar chart (1st visualization)	Stacked bar chart (1st visualization) and Heatmap (3rd visualization)
Engagement (Highest better)	Choropleth map (2nd visualization)	Scatterplot chart (4th visualization)	Stacked bar chart (1st visualization) and Bar Chart (5th visualization)	Stacked bar chart (1st visualization) and Heatmap (3rd visualization)
Usefulness (Highest better)	Heatmap (3rd visualization)	Bar Chart (5th visualization)	Bar Chart (5th visualization)	Stacked bar chart (1st visualization)
Usefulness of components (Highest better)	Stacked bar chart (1st visualization)	Choropleth map (2nd visualization)	Heatmap (3rd visualization)	Stacked bar chart (1st visualization)
Efficacy (Highest better)	Scatterplot chart (4th visualization)	Bar Chart (5th visualization)	All the same besides the Scatterplot chart (4th visualization)	Stacked bar chart (1st visualization)

Average rankings (Highest better)	Heatmap (3rd visualization)	Heatmap (3rd visualization)	Choropleth map (2nd visualization)	Choropleth map (2nd visualization)

Regarding the experts group the worst visualization was also the second visualization, for the same reason as the novice's, since it was also the visualization that ranked in last and since it was one of the visualizations that had one of the lowest average of usefulness, lowest average of efficacy and the highest completion time. It is also possible to consider that the first visualization is also one of the worst for the expert's group, considering it had the lowest averages of engagement and the highest averages of complexity in general and regarding the components, and also the lowest accuracy average. This means, that similar to the novice's group the worst visualizations were also the second and the first visualizations, meaning that no differences were found between the two groups regarding this aspect and that both groups had difficulties with this type of idioms, stacked bar chart and choropleth map.

In **Table 19**, it is also possible to see the most and least useful and complex components for each group. With this analysis, it is possible to conclude that regarding the interactive features, besides the actual visualizations, for the novice's group, the radio box and the text can be one of the most complex components, whereas for the expert's group is the select box, meaning that for the expert's group the radio box may be a more suitable component and for the novice's group it might be the select box. The text, depends on the use, considering it is for both groups sometimes one of the most complex components and one of the least complex. The slider and the tooltips are considered in both groups to be one of the least complex components, which means that both groups are comfortable interacting with these features, and they might be adequate for both groups.

Table 19 - Complexity and usefulness of the components for the different groups

Components	**Novices**	**Experts**
Most Complex	Radio box and text	Select box
Least Complex	Slider and tooltips	Slider and tooltips
Most Useful	Slider	Radio box, the slider and the tooltips
Least Useful	Text	Text

As for the most useful components, both groups considered most of the interactive features to be useful, either the slider, the tooltips and the radio box, but in the novice's group the slider is mentioned in two visualizations, where in the expert's group the radio box, the slider and the tooltips are also mentioned in two visualizations, meaning the expert's group might find the radio box to be more useful than the novice's group, which is coherent with the previous analysis. In conclusion the most useful component for the novice's group is the slider and for the expert's group is the radio box, the slider and the tooltips.

The least useful component, was in both groups and in all of the visualizations the text, meaning that the participants of both groups found that the visualizations and their components were more useful, likely due to the fact that they were interactive and the text was the only component that wasn't, concluding that both groups prefer features that are interactive and associate interactivity with usefulness. It is also worth noticing that, as previously mentioned, the only interactive feature that was not mentioned as one of the most useful components was the select box for the novice's group, which means that although the select box is less complex for the novice's group it might not be as useful, but considering it was not ranked as one of the least useful components and since the complexity is lower than the radio box, the select box might still be more suitable than the radio box for the novice's group.

Comparing this results with the existing literature, using the example in the study conducted by Blascheck, *et al.* (2019), it is possible to see a major difference between this study and the project in question, considering that, although this study also used different types of individuals in their experiences, it did not find any differences between those groups, which in turn, is not true for this project, since there were differences identified in the way both groups perceived the different components/interactive features and regarding the qualitative measure engagement.

When comparing the results of this study, with the results of the study conducted by Oghbaie, Pennock, & Rouse (2016) and Rouse, Pennock, Oghbaie, & Liu (2017) it is possible to observe that there are some differences and similarities. Regarding the completion time, which can be interpreted as the speed, considering that no statistically significant differences were found between both groups regarding the completion time measure it can be concluded that the speed was the same for both groups, as it was in the case of both studies being compared. Regarding the accuracy, the results were different than the results in the work

conducted by Rouse, Pennock, Oghbaie, & Liu (2017), since the experts group did not have any statistically significant differences than the novice's group regarding the accuracy measure, and they were also different than the results in the work conducted by Oghbaie, Pennock, & Rouse (2016), that concluded that the experts only exceed non-experts in the data with the most complex casual relationships.

Lastly, another work that can be compared with the results of this study is the work conducted by Géryk (2015). This work concluded that using animations in visualizations can lead to a decreased number of errors, which can correlate with this study, considering that all of the visualizations in the prototype were interactive and all of them had high values of accuracy, with the lowest value being 87% and the highest 97%. Correspondingly, the average value of wrong answers was 1.67 for the novice's group and 1.87 for the expert's group, meaning that as the work conducted by Géryk (2015) concluded, the interactivity might be associated with fewer errors. Another interesting observation is that the visualization that obtained the lowest accuracy was the only visualization where the idiom was not interactive, which was the second visualization, being coherent with the results of Géryk (2015). Finally, it is also interesting to notice that this work also used qualitative and quantitative measures to validate its visual analytical tool, similar to this study.

Other additional findings that were found in this study were that it was proved that the Shiny package is a tool that is capable of creating interactive visualizations for different types of individuals, since both groups obtained high averages of accuracy in all of the visualizations and in general, and all of the average values of usefulness, efficacy and engagement of all the visualizations and in general ranked in the High interval. Moreover, the average values of complexity of the prototype in general ranked in the Very low interval meaning the visualizations and the prototype in general were easy to analyse. These results can be compared to the work conducted by López *et al.* (2018), considering that the author discovered that some students found the application to be useful and complemented the learning experience, and a lot of the students responded in a positive manner, which can be correlated with the results from this study, considering that on average the participants from this study found the prototype and each of the visualizations to be useful, and in the experiment observations some participants also mentioned that they enjoyed the prototype and that it was a useful tool to learn. Lastly, another thing that is worth mentioning, is that, as mentioned previously, when building the prototype from this study, the recommendations

from the study conducted by López *et al.* (2018) were also followed, which might have helped to increase the usefulness and reduce the complexity of the prototype in general.

Finally, another result that was interesting to analyse, was that regarding the SI cues that were implemented in the prototype, the SI cue that was the most useful for the novice's group was the second SI cue, considering that the most useful component for this group, besides the actual visualization, was the slider, and in that component the SI cue that is present is in an external object. As for the group of expert's, considering that the most useful components, besides the actual visualizations, were the slider, radio box and tooltips and that all three of them were classified as the most useful in two visualizations each, the SI cues that were the most useful were the first and the second SI cues, considering that in these components mentioned, there are SI cues present in the object of interest, for example the tooltips, and there are SI cues present in external objects, for example the slider and the radio box. This can correlate with the study conducted by Boy, Eveillard, Detienne, & Fekete (2016), since the SI cues mentioned in this study were used in the prototype, but different results were obtained considering that regarding the work conducted by Boy, Eveillard, Detienne, & Fekete (2016), it was concluded that the only SI cue that was useful to start the interaction of the participants with the visualizations was the third SI cue, on the contrary of this study, where the third SI cues were present in the text, and that was the component that was on average the least useful component for the participants.

6. CONCLUSIONS

Through the ANOVA single factor test we found that the only metric that had statistically significant differences between both groups was the engagement measure. Regarding the visualizations, both groups agreed that the best visualizations were the heatmap and the bar chart and the worst visualizations were the choropleth map and the stacked bar chart. As for the components, the select box was a better option for the novice's group, while the radio box was the best for the expert's group, and the tooltips and the slider were adequate for both types of individuals. With this study, we conclude that, although there are some similarities in how the different types of individuals perceive the interactive visualizations, there are also differences between the two groups and so it is possible for companies and organizations to use these suggestions and adapt their visualizations for the different types of individuals in order to create visualizations that are effective for different types of audience. We also found that the Shiny package is a powerful tool that makes possible the creation of effective interactive visualizations that are suitable for different types of individuals, which can be extremely useful and may provide them with a competitive advantage.

It is also important to take in regard the limitations of this study, for example, the data could have been richer if it was used a sample of a bigger size, or for example if there were created more groups to divide the participants of this study. In addition, another limitation was that some suggestions that were recommend by the participants were not possible to be implemented, and that might have not been the case if there was more experience with the programming language, considering R is considered to be a flexible tool. Moreover, another limitation could be the fact that the prototype only uses five different types of idioms, but considering the time used for the experiences, it would not be feasible to have more visualizations to test.

As for future work, it would be interesting to try this experience with other different groups, for example following the work conducted by Blascheck, *et al.* (2019), using three different types of individuals: novices, individuals that have no knowledge on the subject in the visualization and that have no knowledge on data visualization; savvy, individuals that have knowledge on data visualization but do not have knowledge on the subject in the visualization and lastly, experts, individuals that have knowledge on the subject in the visualization. It is also recommended to recreate the same prototype with different tools to discover which could be more efficient and compare them, as well as using different

visualization idioms in the prototype. Another suggestion is to use a bigger sample size to understand if the same results would be obtained or if they could change. Lastly, it would be useful to continue to understand why the engagement metric was the only metric with statistically significant differences between both groups and to try to discover methods to decrease this difference and completely close the gap between experts and novices.

BIBLIOGRAPHY

Alexandre, I. (2016). Promoting insight: A case study of how to incorporate interaction in existing data visualizations. *Proceedings of the International Conference on Information Visualisation* (pp. 203–208). Lisbon, Portugal.

Banco de Portugal. (2019). *BP Stat Estatisticas Online*. Retrieved from Banco de Portugal Eurosistema: https://www.bportugal.pt/PAS/sem/src/(S(gg1qej45tcbdlvzut2sbrf3n))/selecAnalise.aspx?Token=B542AB22-FA09-4822-AE4C-CFEF9937B23

Becker, L. & Gould, E. (2019). Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data. *Serials Review*, *45*(3), (pp. 1–5). https://doi.org/10.1080/00987913.2019.1644891

Bhardwaj, P., & Baliyan, N. (2019). Hadoop based Analysis and Visualization of Diabetes Data through Tableau. *2019 Twelfth International Conference on Contemporary Computing (IC3)*, (pp.1–5). Noida, India

Blascheck, T., Vermeulen, L. M., Vermeulen, J., Perin, C., Willett, W., Ertl, T., & Carpendale, S. (2019). Exploration strategies for discovery of interactivity in visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *25*(2) (pp. 1407–1420).

Boy, J., Eveillard, L., Detienne, F., & Fekete, J. D. (2016). Suggested Interactivity: Seeking Perceived Affordances for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1) (pp. 639–648). https://doi.org/10.1109/TVCG.2015.2467201

Diamond, M., & Mattia, A. (2017). Data Visualization: An Exploratory Study into the Software Tools Used by Businesses. *Journal of Instructional Pedagogies*, *18* (pp. 1–7).

Dove, G., & Jones, S. (2012). Narrative Visualization: Sharing Insights into Complex Data. *IADIS International Conference Interfaces and Human Computer Interaction* (pp. 21–23). Lisbon, Portugal.

Ellis, D. A., & Merdian, H. L. (2015). Thinking outside the box: Developing dynamic data visualizations for psychology with Shiny. *Frontiers in Psychology*, *6* (pp. 1–6).

Eurostat. (2018). The Macroeconomic Imbalance Procedure (MIP) introduced. Retrieved from Eurostat Statistics Explained: https://ec.europa.eu/eurostat/statistics-explained/index.php/The_Macroeconomic_Imbalance_Procedure_%28MIP%29_introduced

Fahad, A., & Yahya, A. E. (2018). Big Data Visualization: Allotting by R and Python with GUI Tools. *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, (pp. 1–8). Kuala Lumpur, Malaysia.

Géryk, J. (2015). Using Visual Analytics Tool for Improving Data Comprehension. *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 327–334). Madrid, Spain.

International Monetary Fund. (2009). *Balance of Payments and International Investment Position Manual* . IMF Multimedia Services Division.

Heer, J., Kong, N., & Agrawala, M. (2009). Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations. *ACM Human Factors in Computing Systems (CHI)*, (pp. 1303–1312). New York, USA.

Henry, N., & Fekete, J. (2006). MatrixExplorer: a Dual-Representation System to Explore Social Networks. *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 06)*, *12*(5), (pp. 677–684).

Jiang, Z., & Carter, R. (2018). Visualizing library data interactively: two demonstrations using R language. *Library Hi Tech News*, *35*(5) (pp. 14–17).

Kumatani, S., Itoh, T., Motohashi, Y., Umezu, K., & Takatsuka, M. (2016). Time-varying data visualization using clustered heatmap and dual scatterplots. *Proceedings of the International Conference on Information Visualization*, (pp. 63–68). Lisbon, Portugal. https://doi.org/10.1109/IV.2016.50

López, M., Cobo, E., & Cortés, J. (2018). Assessing Shiny apps through student feedback : Recommendations from a qualitative study. *Computer Applications in Engineering Education*, 26(5), (pp. 1061–3773). https://doi.org/10.1002/cae.21932

Lu, F., Yu, D., Liang, H. N., Chen, W., Papangelis, K., & Ali, N. M. (2018). Evaluating Engagement Level and Analytical Support of Interactive Visualizations in Virtual Reality Environments. *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR*, (pp. 143–152). Munich, Germany.

Mcguffin, M. J. & Balakrishnan, R. (2005). Interactive Visualization of Genealogical Graphs. *IEEE Symposium on Information Visualization (InfoVis)*, (pp. 17–24). Minneapolis, USA.

Mclachlan, P., Munzner, T., Koutsofios, E., & North, S. (2008). LiveRAC: Interactive Visual Exploration of System Management Time-Series Data. *ACM Conf. Human Factors in Computing Systems (CHI)*, (pp. 1483–1492). Florence, Italy.

Microsoft Power BI. (2020). *Turn data into opportunity*. Retrieved from Microsoft: https://powerbi.microsoft.com/en-us/

Mottus, A., Kinshuk, Graf, S., & Chen, N. S. (2013). Visualization and interactivity in the teacher decision support system. *Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013* (pp. 502–503). Beijing.

Munzner, T. (2015). Visualization Analysis and Design. CRC Press.

Noack, A. (2003). An Energy Model for Visual Graph Clustering. *11th International Symposium on Graph Drawing (GD 2003), LNCS, 2912*, (pp. 425–436). Springer-Verlag.

Oates, B. J. (2005). Design and Creation. In Researching Information Systems and Computing (pp. 108–124). Sage Publications.

Oghbaie, M., Pennock, M. J., & Rouse, W. B. (2016). Understanding the Efficacy of Interactive Visualization for Decision Making for Complex Systems. *2016 Annual IEEE Systems Conference (SysCon)* (pp. 1–6). Orlando, FL.

Paul, C. L., Chang, J., Endert, A., Cramer, N., Gillen, D., Hampton, S., … Cook, K. A. (2019). TexTonic: Interactive visualization for exploration and discovery of very large text collections. *Information Visualization*, 18(3), (pp. 339–356). https://doi.org/10.1177/1473871618785390

Pena-Araya, V., Pietriga, E., & Bezerianos, A. (2019). A Comparison of Visualizations for Identifying Correlation over Space and Time. IEEE Transactions on Visualization and Computer Graphics, 26(1), (p. 1–1). https://doi.org/10.1109/tvcg.2019.2934807

Phan, D., Xiao, L., Yeh, R., Hanrahan, P., & Winograd, T. (2005). Flow Map Layout. *IEEE Symposium on Information Visualization (InfoVis)*, (pp. 219–224). Minneapolis, MN.

Rouse, W. B., Pennock, M. J., Oghbaie, M., & Liu, C. (2017). Interactive visualizations for decision support: Application of Rasmussen' s abstraction-aggregation hierarchy. *Applied Ergonomics*, *59*(Part B) (pp. 541–553). https://doi.org/10.1016/j.apergo.2016.03.006

Schenker, J. D., & Rumrill, P. D. (2016). Perspectives on Scientific Inquiry Causal-comparative research designs. Journal of Vocational Rehabilitation, 21(3) (pp. 117–121).

Seal, A., & Wild, D. J. (2018). Netpredictor: R and Shiny package to perform drug-target network analysis and prediction of missing links. *BMC Bioinformatics*, *19*(1) (pp. 1–10).

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), (pp. 1139–1148). https://doi.org/10.4185/RLCS-2014-1021

Su, W., Sun, J., Shimizu, K., & Kadota, K. (2019). TCC-GUI: A Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Research Notes*, *12*(1) (pp.1–6).

Sun, J., & Hsu, Y. (2012). Interacting with Computers An experimental study of learner perceptions of the interactivity of web-based instruction. Interacting with Computers, 24(1) (pp. 35–48). https://doi.org/10.1016/j.intcom.2011.11.001

Viorel, N. C., & Lucia, N. (2019). Analysis of information on tourism in the european union using the power bi business analysis service. *Lucrari Stiintifice, Universitatea de Stiinte Agricole Si Medicina Veterinara a Banatului, Timisoara, Seria I, Management Agricol*, *21*(1), (pp. 175–178).

Zhang, C. (2015). Using Excel's Data Table and Chart Tools Effectively in Finance Courses. Journal of Accounting and Finance, 15(7), (pp. 79–93).

Zhang, Y., & Maciejewski, R. (2017). Quantifying the Visual Impact of Classification Boundaries in Choropleth Maps. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1) (pp. 371–380). https://doi.org/10.1109/TVCG.2016.2598541

Zhu Y. (2007) Measuring Effective Data Visualization. Advances in Visual Computing. ISVC 2007. Lecture Notes in Computer Science, 4842. Springer, Berlin, Heidelberg

APPENDIX

APPENDIX A – PROTOTYPE SOURCE CODE

The source code repository for the original prototype can be accessed at the following link:

https://github.com/m20180646/MasterWorkProject

The source code repository for the prototype after alterations, can be accessed at the following link:

https://github.com/m20180646/MasterWorkProjectAlterations

APPENDIX B – PROTOTYPE LINK

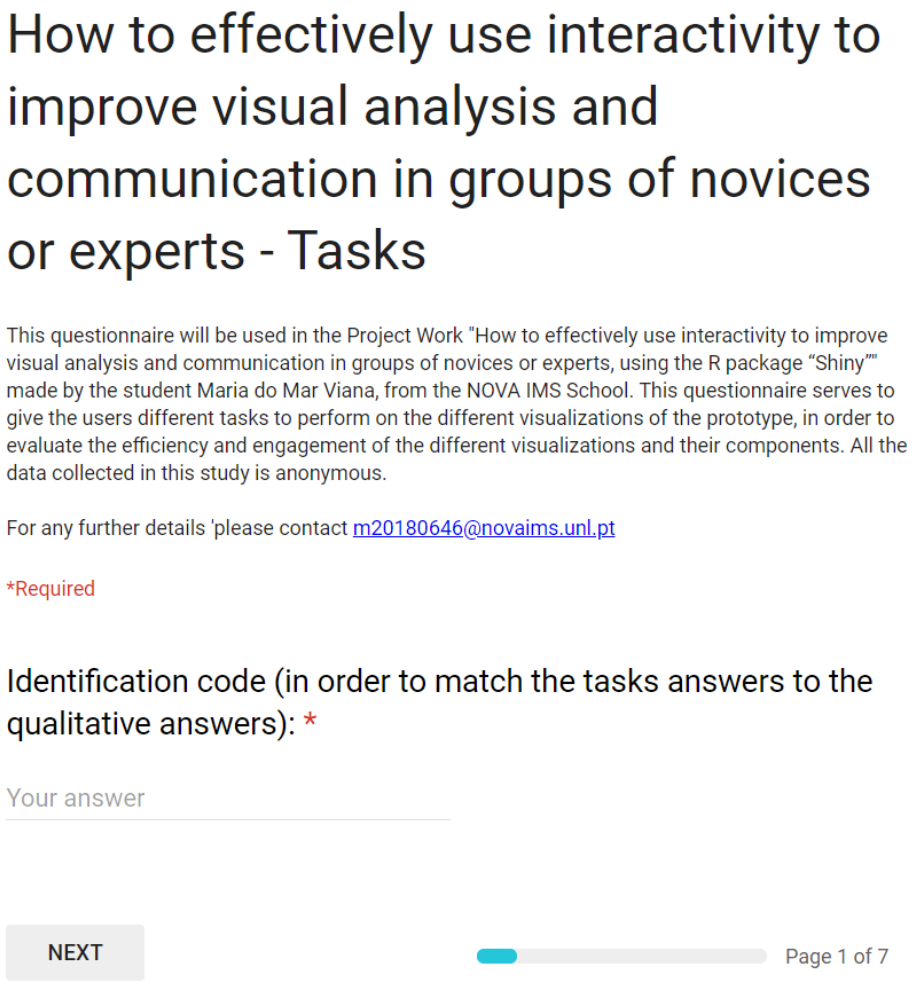
The original prototype, before the implementation of the suggestions of the experiments, can be accessed using the following link:

https://prt-external-transactions.shinyapps.io/datavizapp/

Due to memory constrains regarding the Shiny apps online platform it was not possible to publish the final prototype online, with the suggestions from the experiments.

ANNEXES

ANNEX A – EXPERIMENT TASKS



How to effectively use interactivity to improve visual analysis and communication in groups of novices or experts - Tasks

This questionnaire will be used in the Project Work "How to effectively use interactivity to improve visual analysis and communication in groups of novices or experts, using the R package "Shiny"" made by the student Maria do Mar Viana, from the NOVA IMS School. This questionnaire serves to give the users different tasks to perform on the different visualizations of the prototype, in order to evaluate the efficiency and engagement of the different visualizations and their components. All the data collected in this study is anonymous.

For any further details 'please contact m20180646@novaims.unl.pt

*Required

Identification code (in order to match the tasks answers to the qualitative answers): *

Your answer

NEXT

Page 1 of 7

Figure 25 - Experiment Tasks Questionnaire

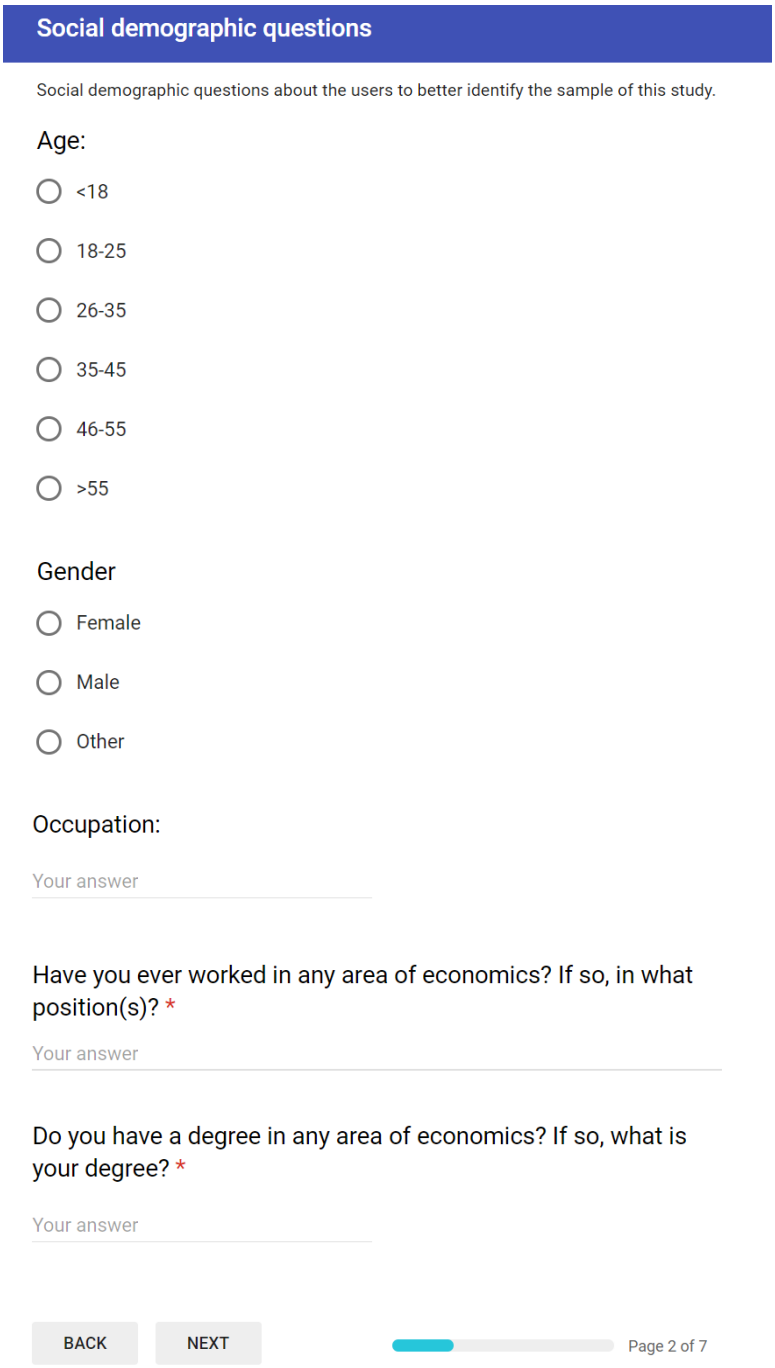


Figure 26 - Social Demographic Questions

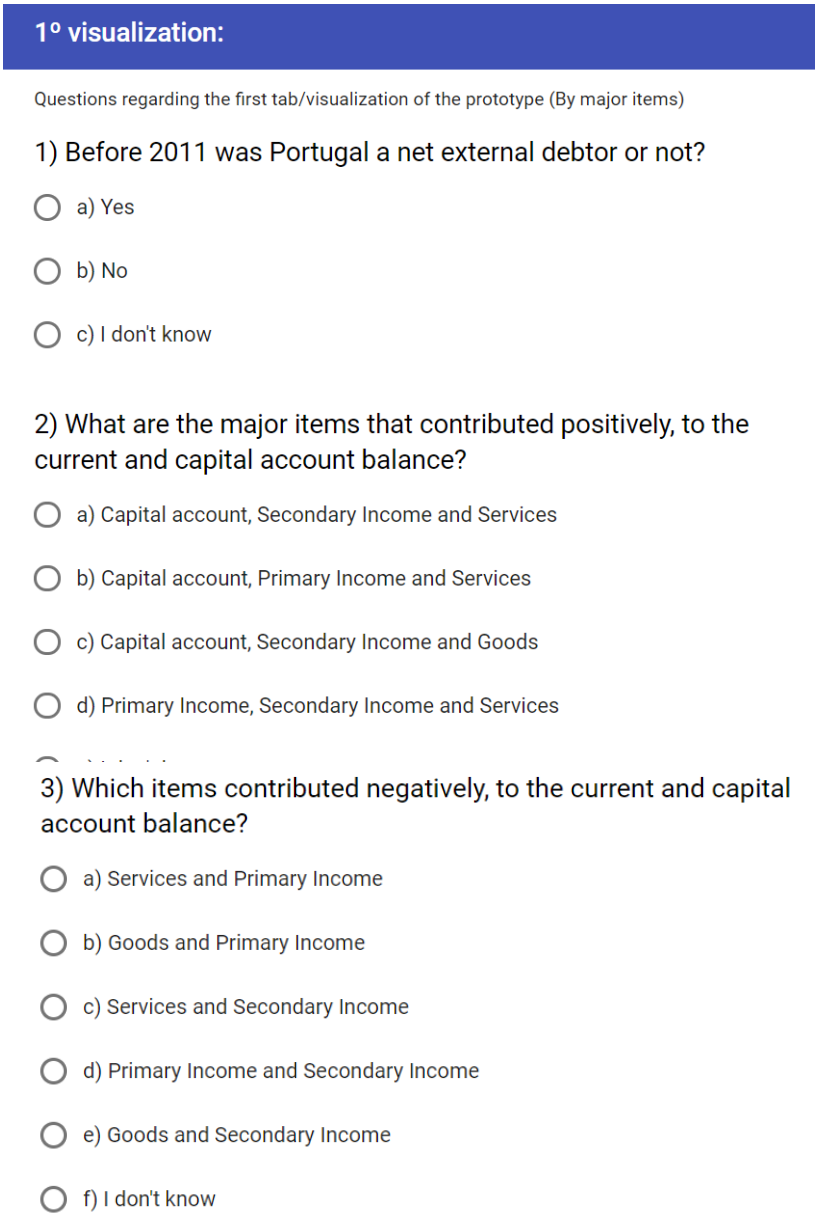


Figure 27 - Tasks Questions 1 to 3 “By major items”

4) In the year 2013, what does the peak of the current and capital account means?

- a) It means that the sum of the Capital account, Secondary Income and Services items is superior to the sum of Goods and Primary Income items.
- b) It means that the sum of the Goods and Primary Income items is superior to the sum of Capital account, Secondary Income and Services items.
- c) It means that the Portuguese economy is borrowing from the rest of the world
- d) It means that the Portuguese economy is lending to the rest of the world
- e) Both a) and d)
- f) Both b) and c)
- g) I don't know

5) Which year has the most negative value of the current and capital account and which item influenced this value the most?

- a) 2008, Services
- b) 2010, Services
- c) 2008, Goods
- d) 2010, Goods
- e) I don't know

6) In the year 2002, what is the specific value of the Goods item?

- a) – 17 376,54
- b) – 1 962,36
- c) – 15 630,56
- d) – 7 269,44
- e) I don't know

BACK NEXT Page 3 of 7

Figure 28 - Tasks Questions 4 to 6 "By major items"

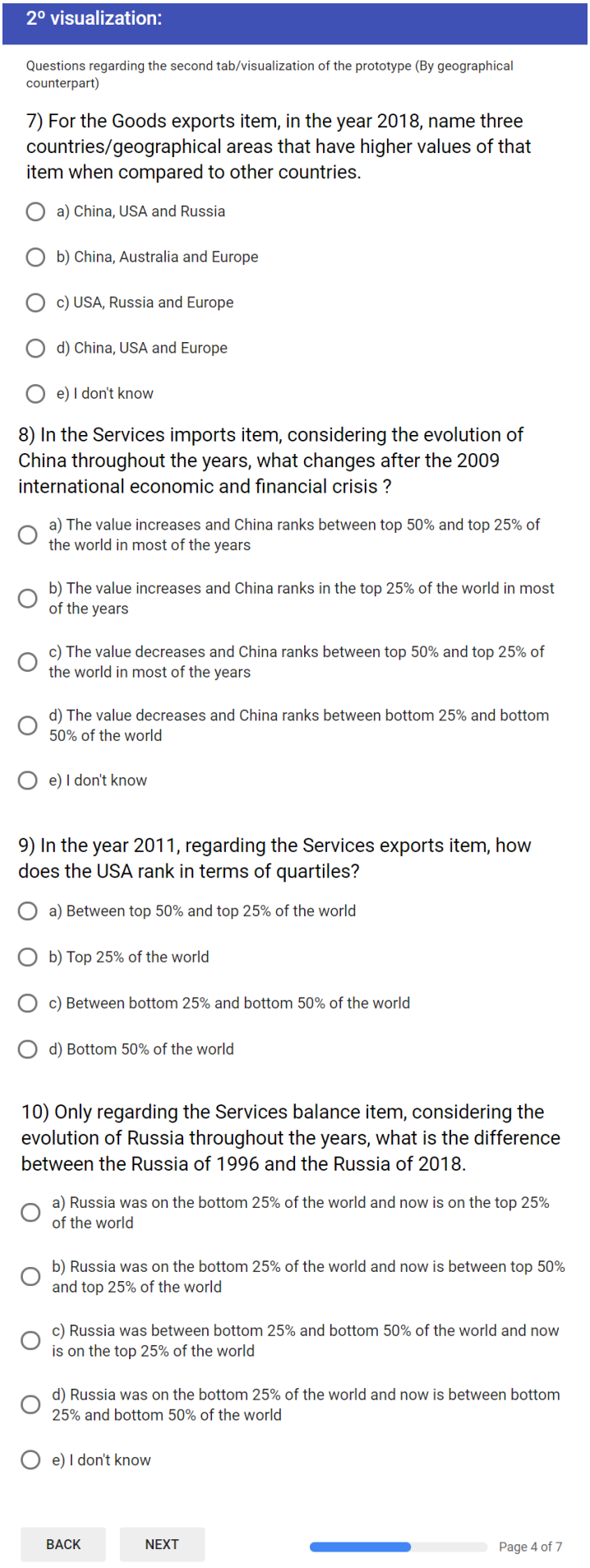


Figure 29 - Tasks Questions 7 to 10 "By geographical counterpart"

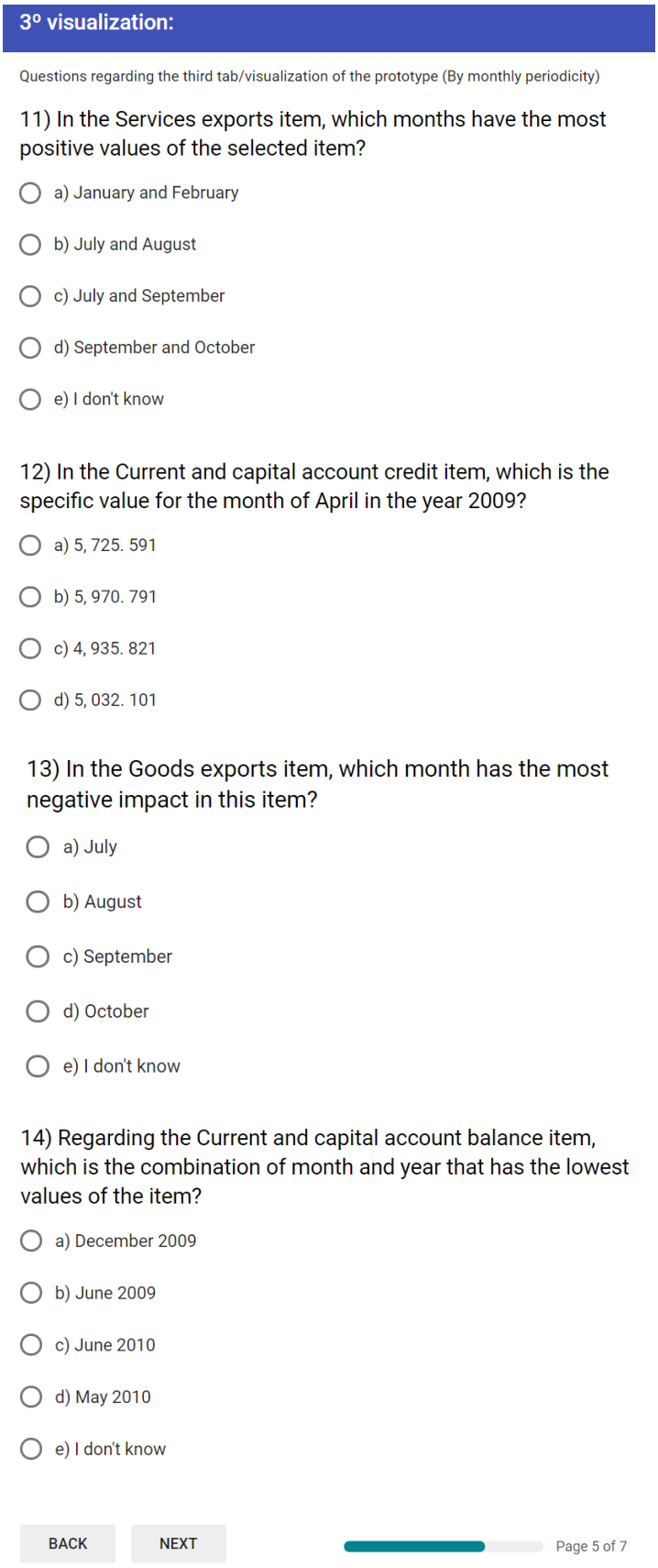


Figure 30 - Tasks Questions 11 to 14 "By monthly periodicity"

4º visualization:

Questions regarding the fourth tab/visualization of the prototype (By type of services)

15) Which type of services has the most impact on the global value, throughout the years and regarding all services?

- a) Government and Transport
- b) Transport and Travel
- c) Financial Services and Transport
- d) Financial Services and Travel
- e) I don't know

16) When selecting the option to exclude the Transport and Travel services, in the year 2018, which type of service has the most impact on the global value?

- a) Telecommunications and computers
- b) Intellectual property
- c) Government
- d) Other business services
- e) I don't know

17) What does it mean when a service is located above the red line (bissector) ?

- a) It means that the service has values of exports greater than the imports
- b) It means that the service has values of exports lower than the imports
- c) It means that the service has values of exports equal to the value of imports
- d) I don't know

Figure 31 - Tasks Questions 15 to 17 "By types of services"

18) What is the percentage of impact that the service Telecommunications and computers has on all the services excluding Travel and Transport, in the year 2009?

- a) 2 %
- b) 4.1 %
- c) 4.44 %
- d) 3 %
- e) I don't know

19) What is the value of the exports for the Travel service in the year 2012?

- a) 5152. 59
- b) 8605.54
- c) 9249. 61
- d) 5574. 71
- e) I don't know

BACK NEXT Page 6 of 7

Figure 32 - Tasks Questions 18 and 19 "By types of services"

5º visualization:

Questions regarding the fifth tab/visualization of the prototype (By international comparison)

20) Throughout the years, comparing Portugal with all European countries, when does Portugal start to have positive values of current account balance?

- a) 2012
- b) 2010
- c) 2015
- d) 2017
- e) I don't know

21) What is the specific value of the current account balance for Spain, in the year 2011?

- a) – 2.399
- b) – 2.199
- c) – 8.899
- d) – 3.799
- e) I don't know

22) When comparing Portugal with Italy, Greece and Spain, where does Portugal rank in 2017?

- a) 1st
- b) 2nd
- c) 3rd
- d) 4th
- e) I don't know

Figure 33 - Tasks Questions 20 to 22 "By international comparison"

23) Using the option to compare Portugal with all of the countries that entered the EU after 2004, where does Portugal rank in 2015?

- a) 4th
- b) 8th
- c) 6th
- d) 1st
- e) I don't know

24) When comparing Portugal to all the European countries, which country has the highest value of current account balance in 2014?

- a) Denmark
- b) Netherlands
- c) Spain
- d) Sweden
- e) I don't know

BACK SUBMIT Page 7 of 7

Figure 34 - Tasks Questions 23 and 24 "By international comparison"

ANNEX B – QUALITATIVE QUESTIONS

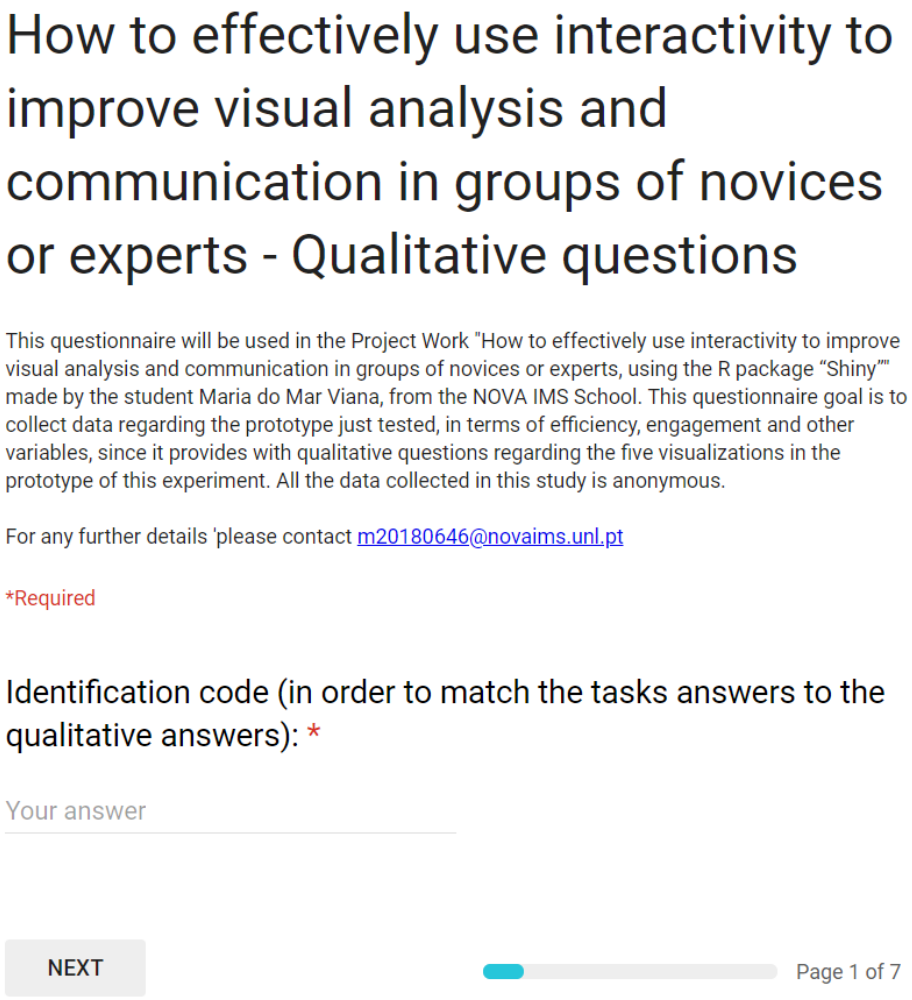


Figure 35 - Qualitative Questions Questionnaire

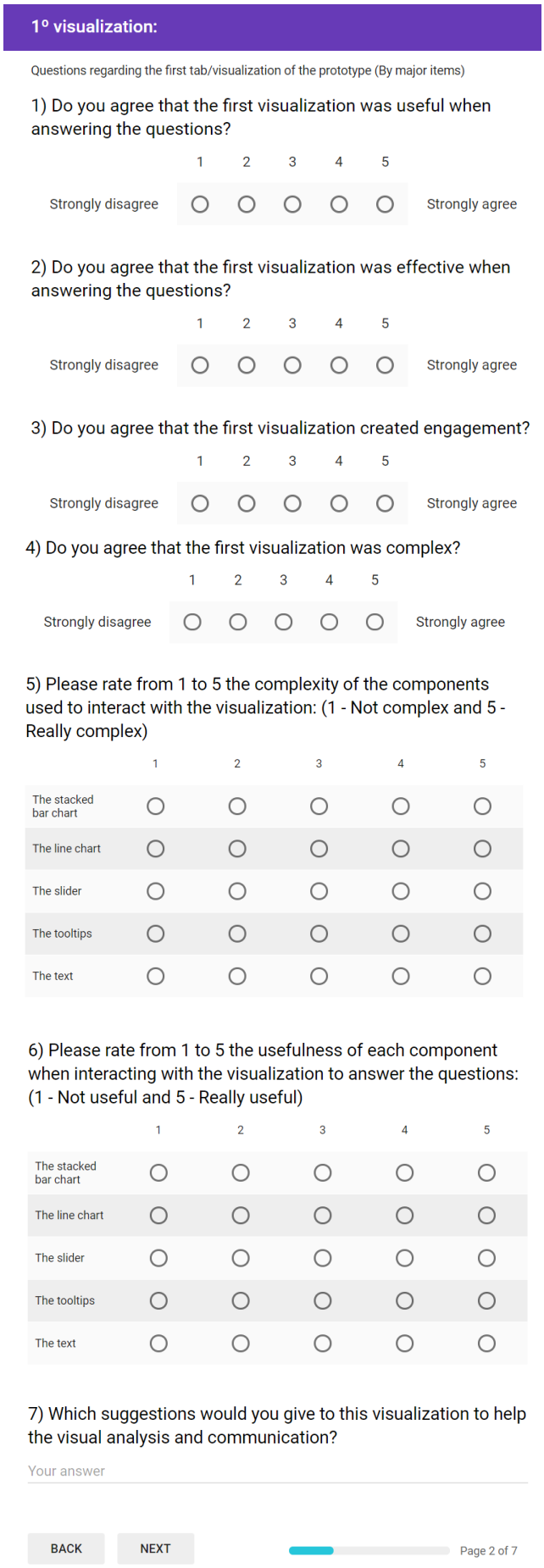


Figure 36 - Qualitative Questions "By major items"

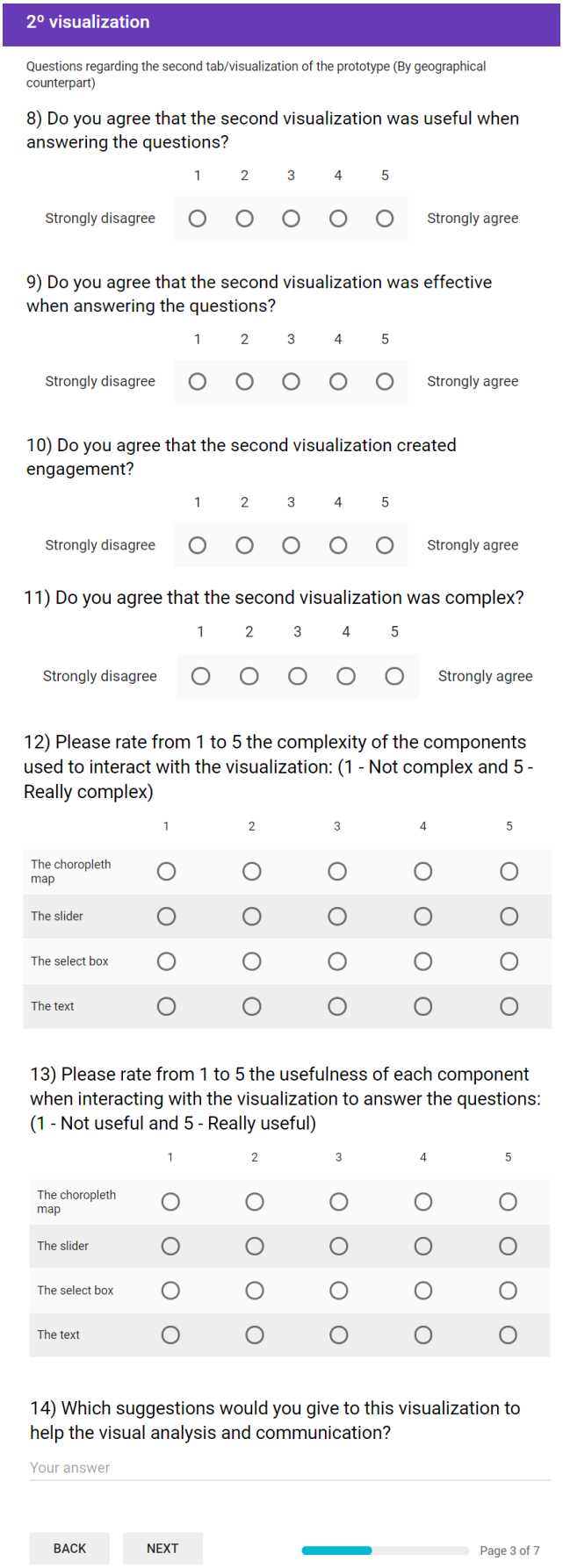


Figure 37 - Qualitative Questions "By geographical counterpart"

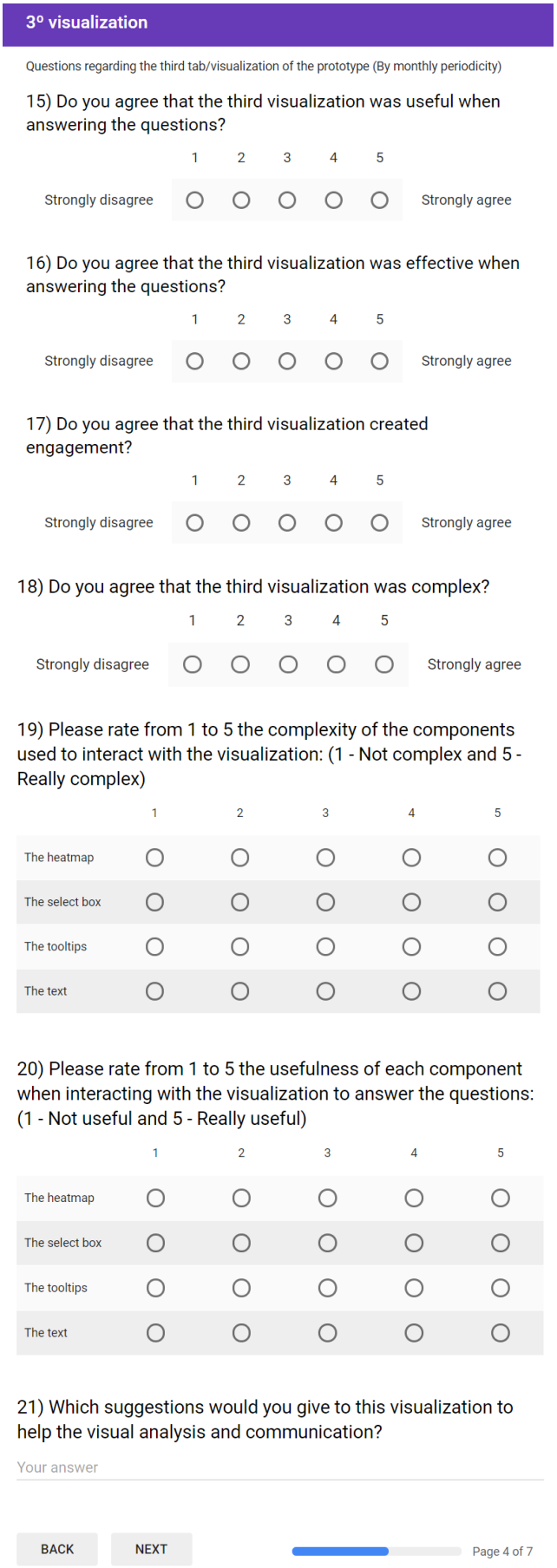


Figure 38 - Qualitative Questions "By monthly periodicity"

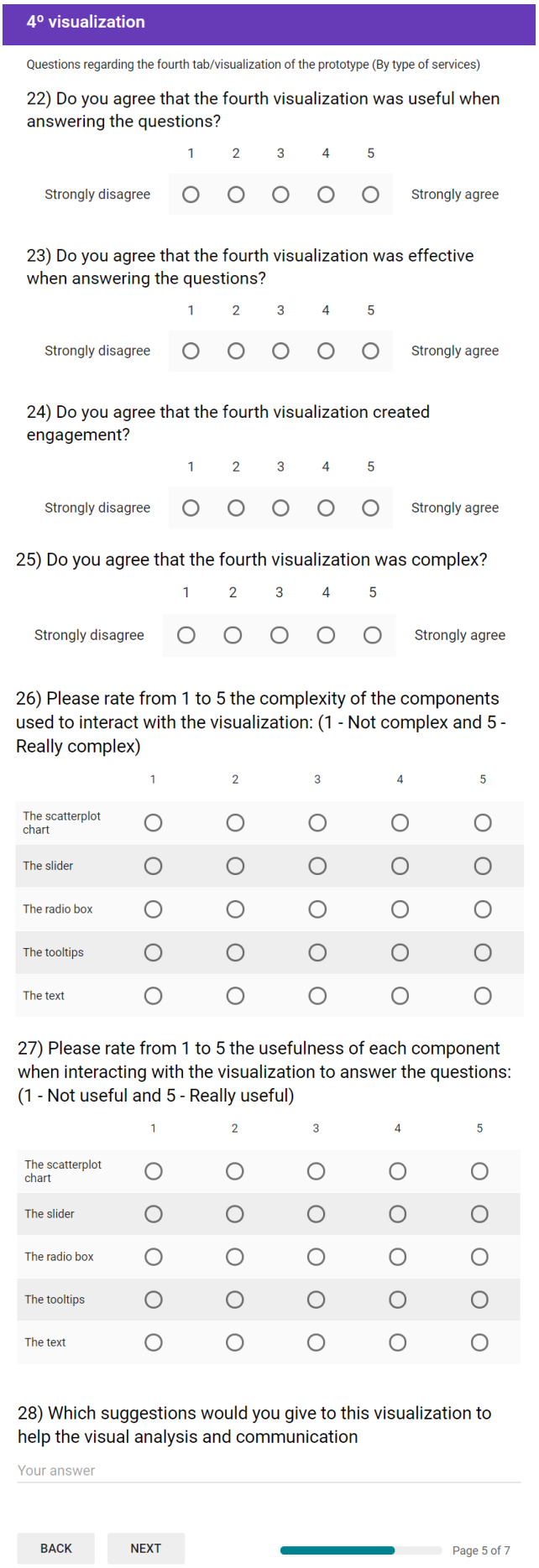


Figure 39 - Qualitative Questions "By type of services"

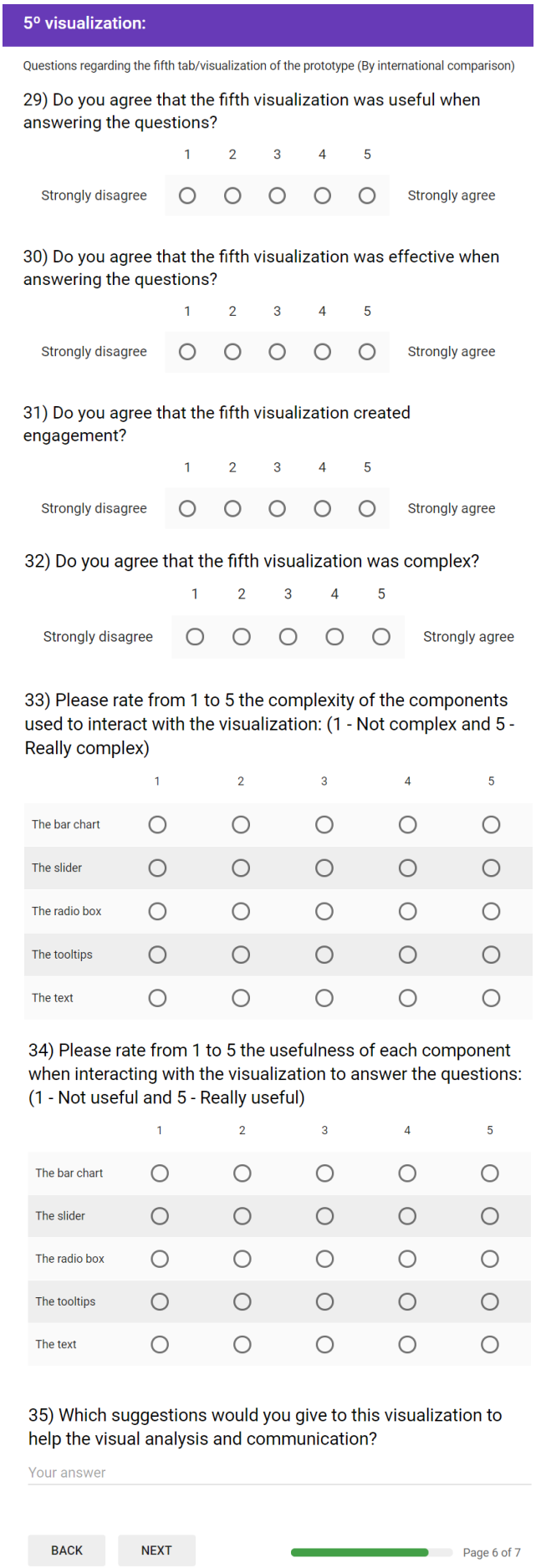


Figure 40 - Qualitative Questions "By international comparison"

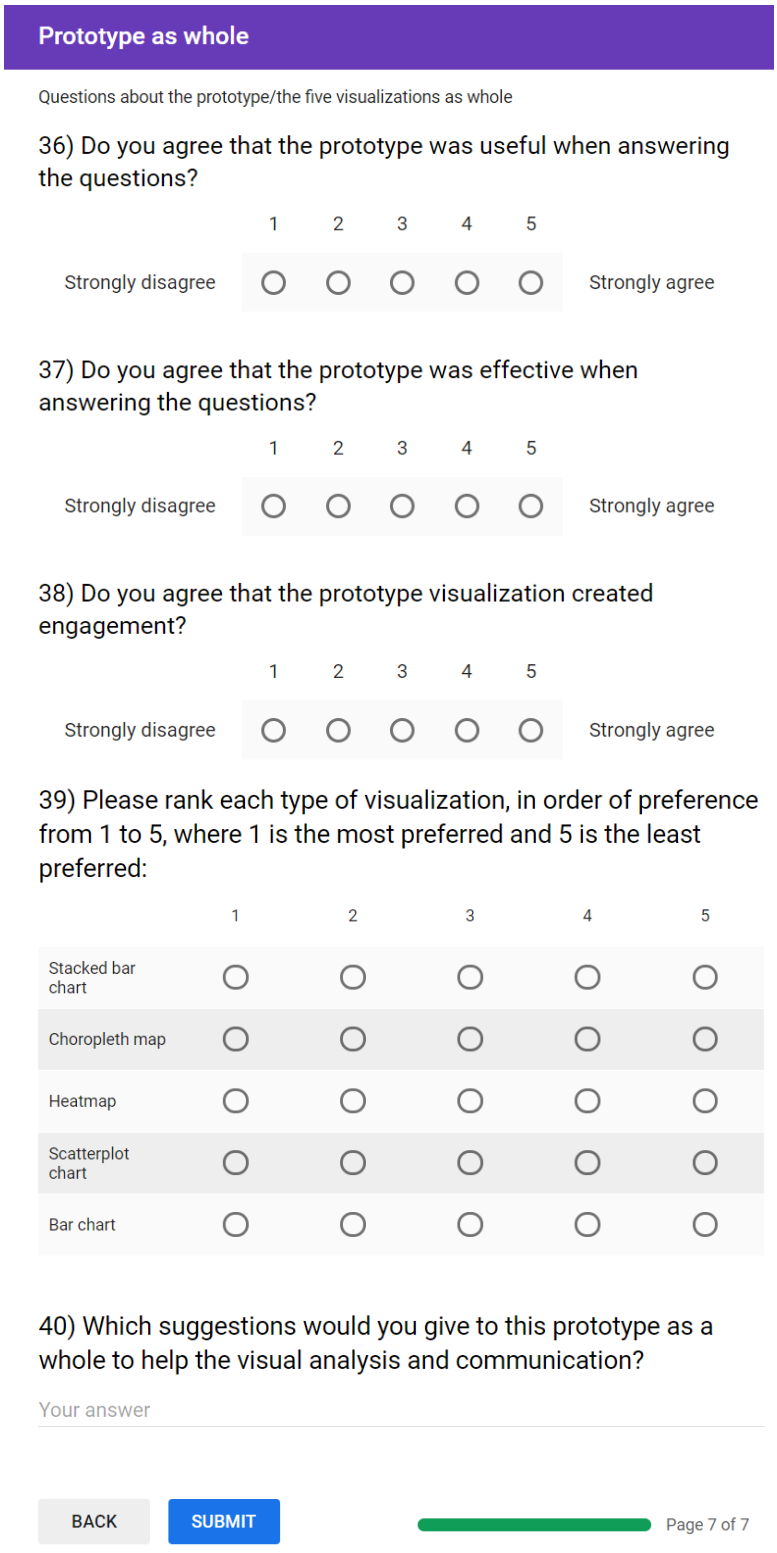


Figure 41 - Qualitative Questions “Prototype as a whole”

ANNEX C – SUGGESTIONS MADE IN THE EXPERIMENTS

Table 20 - Suggestions made in the experiments

Suggestions	Novices	Experts	Both groups	I*
The play button could be more prominent.	3	1	X	Y
The play button could be in a lower position.	5	1	X	N (R1)
The map of the second visualization could have values and a legend for each country.	4	2	X	N (R2)
Add the total variable (goods + services) in the select box of the second visualization.	-	1	-	N (R3)
Wanted the animation of the slider to be slower.	1	1	X	Y
In the first visualization the order of the legend should be ordered by the order of the items in the graph (by the colours in the graph).	-	1	-	N (R1)
The heatmap could have less colour variance to improve visibility.	1	1	X	Y
The size of the letters could be bigger.	1	2	X	N (R2)
In the first visualization the colours could be different. For example, the most logical colour for the primary income should be green.	-	1	-	Y
In the second visualization the map could have a zoom and could be aggregated for economical geographical areas.	1	1	X	N (R2)
The third visualization is different than others because it doesn't have a slider and the time analysis is more difficult. The same logic should be applied to all visualizations.	-	2	-	N (R2)
The radio box variables could be more highlighted.	1	-	-	Y
In the fifth visualization the ISO Codes 2 could be used to facilitate the analysis.	1	1	X	Y
It should be possible to export the values of the second visualization as it is in the others.	-	1	-	N (R1)

The first visualization had too much text and that the text could be presented in a different format.	-	2	-	N (R2)
The format of the tooltips should be more constant and more logic.	2	3	X	Y
The slider in the first visualization should not allow for only one year to be selected.	2	-	-	N (R1)
The colours in the scatterplot of the fourth visualization could be misleading (the blues are very similar) so it would be best to change them.	1	-	-	Y
Use different colours in the map of the second visualization, maybe red or orange to be more effective in distinguishing.	-	1	-	Y
Use more distinguishable colours between Portugal and the other countries in the fifth visualization (because red and yellow are of the same temperature).	-	1	-	Y
In the heatmap it would be nice to click on the values and have a drill down for a better analysis.	-	1	-	N (R1)
In the first visualization there could be a first seasonal analysis (most relevant years), and then a drill down for a yearly analysis.	-	1	-	N (R3)
In the third visualization the select box could have a breakdown in the options. Have a radio box and then a select box inside.	-	1	-	N (R1)
In the fourth visualization highlight the travel and transport without removing the colours, for example using textures or shades in the actual colours.	-	1	-	Y
In the fifth visualization Portugal could be highlighted in the label as well, so it can be visible even if it has null values.	-	2	-	N (R1)
In the fifth visualization, more choices could be added in the radio box, for example compare it with the Nordic countries, or add a productivity variable such as the fiscal taxes that a country has, to create a better and richer analysis.	-	1	-	N (R3)
Implement a tab with a network.	-	1	-	N (R2)
Have a reset button for all the visualizations.	-	1	-	Y

It would be simpler in the cases of specific years to just have a box to enter the year wanted, instead of selecting a specific year in the slider (without the play) and only allow the use of the slider with play.	2	-	-	N (R1)
In the fifth visualization when touching the ticks/labels in the graph the full name of the country could appear.	-	1	-	Y
Put in the text that the data from the second visualization is related to Portugal.	-	1	-	Y
To have ticks in the slider so it would be easier to select a specific year.	-	1	-	Y
Underline or put the text in the second visualization about the quartiles in bold.	-	1	-	Y
The size of the heatmap could be increased to facilitate analysis, because it is easy to do mistakes.	1	-	-	Y
Mention in the text of the fourth visualization what does the variable weight mean, to facilitate the analysis.	2	-	-	Y
In the fifth visualization, if possible, thinks the graph should be stretched so it would be easier to see the values of Portugal.	1	-	-	Y
Increase the size of the sidebar or to mention in the text that there are two options in the radio box in order to highlight that, in the fourth visualization.	1	-	-	Y
In the third type of visualization organize the items in the select box by all goods, all services and all current and capital account instead of organizing by all exports, all imports or all balances.	1	1	X	Y
In the fifth visualization show the full name of the countries in the graph, instead of showing the ISO codes.	1	-	-	N (R2)
Suggested a + and a – to select the specific years in the slider of the first visualization.	-	1	-	N (R1)
The slider bar should not be filled with colour because it is not a cumulative analysis.	-	1	-	N (R1)
In the first visualization the stacked bars could have spaces between each component to better understand the graph and improve the analysis.	-	1	-	N (R1)

The select box could be in a lower position in the third visualization.	1	-	-	Y
In the first visualization the second part of the text about each item should be under the graph and the size of the graph could be increased.	1	-	-	N (R2)
In the second visualization there could be a table with all the real values of each country under the map.	1	-	-	N (R2)
Suggested to put an annotation in the fourth visualization to explain the bissectrice or to underline or put in bold the section that explains that in the text.	-	1	-	Y
Put the flags in the bars of the fifth visualization to better identify the countries.	-	1	-	N (R1)
In the fifth visualization, thinks a table or a line chart could be better to represent the data, if the idea was to highlight Portugal.	-	1	-	N (R2)
The y axe of the fifth visualization could be dynamic and adapt to the values.	-	1	-	Y
In the second visualization it should explain in the text the meaning of quartiles.	-	1	-	Y
The play button would look better in the beginning of the slider bar.	1	-	-	N (R1)
Restructure the text in the fifth visualization.	-	1	-	Y

*In **Table 20**, it was necessary to resort to the use of the following acronyms:

I Implemented

N No

R1 Reason 1

R2 Reason 2

R3 Reason 3

Y Yes

ANNEX D – OBSERVATIONS MADE IN THE EXPERIMENTS

Table 21 - Observations made in the experiments

Observations	Novices	Experts	Both groups
The select box was more useful than the radio box.	4	2	X
Really enjoyed the slider.	2	-	-
The radio box is more intuitive than the select box because it is possible to see all the choices.	7	5	X
Only used the tooltips in the task phase.	2	-	-
The play button was very intuitive and engaging.	8	11	X
Had trouble using the slider or the select box.	-	1	-
Had trouble starting to interact.	-	2	-
Had trouble in finding the radio box in the fourth visualization.	-	1	-
Had more trouble interacting with external interactions than with internal interactions such as tooltips.	-	1	-
The radio box is more intuitive than the select box but only if it has few categories. If it has a lot of categories the select box is preferred.	4	6	X
Thought the slider without the play button was more intuitive.	1	2	X
Thought that the fourth visualization was really engaging.	-	1	-
In the first visualization the slider with an interval was more complex and it was hard to select only one year in the slider.	-	1	-
Did not use the play in any of the visualizations.	1	1	X
Preferred to use the slider with the play button.	4	8	X
The first visualization was the most complicated because of the stacked bar chart, but not because of the visualization but because of the topic itself considering the participant was not an expert.	1	-	-
The first visualization was more complex but maybe it was because it was the first impression and because it has a lot of information.	1	1	X
The visualizations were adequate to the data.	-	1	-
The third visualization can be more complex because the select box has a lot of options.	-	1	-

The second visualization with the map was the least intuitive.	-	1	-
In general, it was a good visualization and useful to learn at least the basic and it was very intuitive.	-	2	-
Had trouble analysing the first graph. Had trouble finding the tooltips and interacting with them.	-	1	-
Preferred the slider without the play button if it was for a specific year but to show changes and in general preferred the slider with the play.	3	2	X
In a lot of visualizations, the text was not very useful, with the exception of 1 and 2.	-	1	-
The most difficult visualization was the heatmap because at the first impact it seems very complex and because of all the colours but then once it is understood is not that complex.	-	1	-
The map in the second visualization was the favourite visualization.	-	1	-
The hardest one to analyse was the map but because the user may not know the location of certain countries in the map, and the scatterplot was the hardest because of the visualization itself.	1	-	-
Used the radio box in the fifth visualization to facilitate the analysis.	3	-	-
In the map of the second visualization you can’t really see in the balance items if the values are extremely positive or extremely negative.	-	1	-
The scatterplot chart was the hardest to analyse because it has a lot of elements and the fact that the line in the chart moved was complex.	-	3	-
The map was the simplest one and because of that the participant disliked that one the most.	-	1	-
Did not understand the line in the graph of the four visualization.	2	-	-
The heatmap is more complex because the scales are always changing.	-	1	-
In the fifth visualization is hard to compare two years.	-	1	-
Thinks the scatterplot is the simplest one.	1	-	-
The texts are useful and good at explaining.	1	-	-

Does not like sliders in general, would prefer a combo box but understands is more effective at least in the first visualization.	1	-	-
Thinks the heatmap is good for a macro analysis but more complex for specific cases.	1	-	-
In the fourth visualization clicked on the services in the legend to make them disappear and made several observations about the weight of the transport and travel service and how they were outliers.	1	-	-
The heatmap was the easiest to analyse and the favourite.	3	2	X
The heatmap is not that intuitive and is the least favourite, but because of the visualization type.	-	1	-
In the fifth visualization it may be hard to follow a specific country evolution, that is not Portugal, because the order of the countries changes throughout the years considering the values.	-	1	-
Really liked the bissectrice in the fourth visualization.	-	1	
Had problems in the second visualization to realize if the slider was showing more than one year.	-	1	-
Only used the text in the test phase.	1	2	X